

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 25, 2002, 19:19:18 ; Search time 67 Seconds  
(without alignments)  
795.526 Million cell updates/sec

Title: US-09-889-746-2

Perfect score: 2047  
Sequence: 1 MKSHYFQWLSPLLSAVT.....SKAEVLGRVNPINKPRALF 400

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Database : A\_Geneseq\_101002.\*

1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2047	100.0	400	21	AA07761
2	527	25.7	83	22	AAU33589
3	191.5	9.4	75	22	AAU72914
4	180.5	8.8	78	20	AAU34501
5	180.5	8.8	72	20	AAU34374
6	178.5	8.7	75	20	AAU34472
7	178.5	8.7	76	20	AAU34348
8	171.5	8.4	45	22	ABG17860
9	149	7.3	60	22	ABBS2669
10	147.5	7.2	71	22	ABG17918

11	147.5	7.2	970	14	AAU42451
12	147.5	7.2	970	17	AAU99460
13	147.5	7.2	970	19	AAU37048
14	147.5	7.2	970	20	AAU89398
15	147.5	7.2	970	21	AAU10465
16	146.5	7.2	722	21	AAU07697
17	146	7.1	708	21	AAU75566
18	144.5	7.1	226	22	ABG17885
19	141	6.9	663	22	ABBS2462
20	139.5	6.8	691	21	AAU07698
21	138	6.7	1074	20	AAU43381
22	135.5	6.6	696	17	AAU75366
23	135	6.6	905	22	ABG24651
24	135	6.6	908	22	ABG28872
25	134	6.5	815	22	AAU33610
26	134	6.5	1074	18	AAU35312
27	133.5	6.5	228	22	AAU87504
28	133.5	6.5	228	22	AAU17558
29	133.5	6.5	252	22	AAU87192
30	133.5	6.5	274	22	AAU17142
31	133	6.5	715	17	AAU01461
32	131	6.4	767	19	AAU98606
33	130.5	6.4	793	19	AAU98381
34	130.5	6.4	863	21	AAU87648
35	130.5	6.4	877	19	AAU98214
36	130	6.4	1070	18	AAU35314
37	129.5	6.3	930	18	AAU18061
38	128	6.3	790	17	AAU95568
39	127	6.2	848	20	AAU34512
40	127	6.2	878	20	AAU34385
41	126	6.2	654	22	ABBS2577
42	125.5	6.1	725	20	AAU38832
43	125	6.1	719	17	AAU01462
44	123.5	6.0	529	18	AAU55247
45	123.5	6.0	533	18	AAU55476

## ALIGNMENTS

RESULT 1  
AAB07761  
ID AAB07761 standard; Protein; 400 AA.  
XX  
AC AAB07761;  
DT 07-NOV-2000 (first entry)  
XX  
DE Amino acid sequence of a BASB047 polypeptide.  
XX  
KW BASB047; BASB054; BASB068; BASB069; vaccine; bacteremia; meningitis;  
KW Neisseria meningitidis disease; upper respiratory tract infection;  
KW invasive bacterial diseasee.  
XX  
OS Neisseria meningitidis.  
XX  
PN WO200043519-A2.  
XX  
PD 27-UTL-2000.  
XX  
PF 19-JAN-2000; 2000WO-BP00428.  
XX  
PR 22-JAN-1999; 99GB-0001368.  
PR 28-JAN-1999; 99GB-0001944.  
PR 29-JAN-1999; 99GB-0002086.  
PR 15-FEB-1999; 99GB-0003417.  
PR 16-FEB-1999; 99GB-0003535.  
XX  
PA (SMK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX  
PI Ruelle J;  
XX  
DR WPI; 2000-505839/45.

Enzyme involved in  
Biosynthetic enzym  
S. putrefaciens EP  
S. putrefaciens PK  
Shewanella putrefa  
A Neisseria mening  
Neisseria meningit  
Novel human diagno  
Escherichia coli p  
A Neisseria mening  
M. catarrhalis str  
Adhesin. Escheric  
Novel human diagno  
Novel human diagno  
Pseudomonas aerugi  
M. catarrhalis 422  
Novel central nerv  
Novel signal trans  
Novel central nerv  
Novel signal trans  
Novel central nerv  
H. pylori GHPO 147  
H. pylori GHPO 282  
H. pylori outer me  
H. pylori GHPO 108  
M. catarrhalis Q8  
Pasteurella haemol  
N. gonorrhoeae B h  
Porphyromonas ging  
Porphyromonas ging  
Escherichia coli p  
Neisseria meningit  
NTHI Hxuc protein.  
H. pylori ORF 056p  
H. pylori ORF 07ap

DR N-PSDB; AAS59347.  
XX Neisseria meningitidis BASB047, BASB054, BASB068, and BASB069 proteins,  
PT useful for treating N. meningitidis infections, bacteremia, and  
PT meningitis -  
XX  
XX Claim 3; Page 93-94; 103pp; English.  
XX  
XX The specification describes Neisseria meningitidis polypeptides  
CC designated BASB047, BASB054, BASB068, and BASB069. The polynucleotide  
CC sequences can be used to create a vector to transform a host cell.  
CC The host cell can be used to produce the polypeptide. The polynucleotides  
CC and polypeptides can be used in vaccine compositions. The  
CC polynucleotides, polypeptides, and antibodies directed against the  
CC polypeptides can be used in compositions for preparation of medicaments.  
CC The antibodies can also be used in a composition for treating humans  
CC with Neisseria meningitidis disease. The diseases that can be treated  
CC include upper respiratory tract infection, and invasive bacterial  
CC diseases such as bacteremia and meningitis. The nucleic acid sequences  
CC can be used as probes in the diagnosis of Neisseria meningitidis disease.  
CC The present sequence represents a BASB047 polypeptide.  
XX  
SQ Sequence 400 AA;

Query Match 100.0%; Score 2047; DB 21; Length 400;  
Best Local Similarity 100.0%; Pred. No. 1.6e-161;  
Matches 400; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MRHSYFQWLSPLLSVAVTQOLYAQNESLPTVELEPVVITIDKSGMALANRITQMPHT 60  
Db 1 MRHSYFQWLSPLLSVAVTQOLYAQNESLPTVELEPVVITIDKSGMALANRITQMPHT 60  
QY 61 TKVIEEIQEQATGSRQLADVMAQLIPSLGVSSTTSNFGQTMHGRQVQFLNGVPLTG 120  
Db 61 TKVIEEIQEQATGSRQLADVMAQLIPSLGVSSTTSNFGQTMHGRQVQFLNGVPLTG 120  
QY 121 SRDISRQLNSINPNQVARIIVLSGATSIYGSATGGLINIVTKSDLEEEQETRIGVHGS 180  
Db 121 SRDISRQLNSINPNQVARIIVLSGATSIYGSATGGLINIVTKSDLEEEQETRIGVHGS 180  
QY 181 KLSSEGIGVQVGSVAGVSENGVNLARDVDYRTTGGAFDANGKRIAPPAQTDKQDSKS 240  
Db 181 KLSSEGIGVQVGSVAGVSENGVNLARDVDYRTTGGAFDANGKRIAPPAQTDKQDSKS 240  
QY 241 LSVNTNVDMQLDDKQINLALTHYNDKQDTPDYAPDYGNNRLAVLFGKPSLNAIKGLSLSE 300  
Db 241 LSVNTNVDMQLDDKQINLALTHYNDKQDTPDYAPDYGNNRLAVLFGKPSLNAIKGLSLSE 300  
QY 301 QPKTKSTFNINHHDDLWGNTINTNAYRREKGRFPFVAPFSTAKALPILQSNMLPSA 360  
Db 301 QPKTKSTFNINHHDDLWGNTINTNAYRREKGRFPFVAPFSTAKALPILQSNMLPSA 360  
QY 361 TLDAYTKAPQARAYGVLOSSESAAEVLGRVNLNKKRALF 400  
Db 361 TLDAYTKAPQARAYGVLOSSESAAEVLGRVNLNKKRALF 400

RESULT 2  
AAU33589  
ID AAU33589 standard; Protein; 813 AA.  
XX  
AC AAU33589;  
XX  
DT 14-FEB-2002 (first entry)  
XX  
XX Pseudomonas aeruginosa cellular proliferation protein #33.  
XX  
KW Antisense; prokaryotic cellular proliferation protein;  
KW antibiotic; antibacterial; drug design.  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN WO200170955-A2.

XX PD 27-SEP-2001.  
XX PF 21-MAR-2001; 2001WO-US09180.  
XX PR 21-MAR-2000; 2000US-191078P.  
XX PR 23-MAY-2000; 2000US-206848P.  
XX PR 26-MAY-2000; 2000US-207727P.  
XX PR 23-OCT-2000; 2000US-242578P.  
XX PR 27-NOV-2000; 2000US-253625P.  
XX PR 22-DEC-2000; 2000US-257931P.  
XX PR 16-FEB-2001; 2001US-269308P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Haselbeck R, Ohlsen KL, Zysek JW, Wall D, Trawick JD, Carr GJ;  
XX PI Yamamoto RT, Xu HH;  
XX DR WPI; 2001-611495/70.  
XX DR N-PSDB; AAS51448.  
XX  
PT New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
XX  
XX Example 3; Seq ID No 5085; 511pp; English.  
XX  
XX The invention relates to antisense inhibitors of genes essential to  
XX prokaryotic cellular proliferation, their use in identifying the  
XX genes, their use in the discovery of novel antibiotics, the essential  
XX genes themselves and the encoded proteins. The prokaryotes used are  
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
XX invention is also useful for the identification of potential new targets  
XX for antibiotic development. The antisense nucleic acids can also be used  
XX to identify proteins used in proliferation, to express these proteins,  
XX and to obtain antibodies capable of binding to the expressed proteins.  
XX The proteins can be used to screen compounds in rational drug discovery  
XX programmes. The antisense nucleic acid sequence is also useful to screen  
XX for homologous nucleic acids which are required for cell proliferation in  
XX a wide variety of organisms. The present sequence represents an  
XX essential prokaryotic cellular proliferation protein.  
XX Note: The sequence data for this patent did not form part  
XX of the printed specification, but was obtained in electronic  
XX format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences.  
SQ Sequence 813 AA;

Query Match 25.7%; Score 527; DB 22; Length 813;  
Best Local Similarity 34.2%; Pred. No. 6.1e-35;  
Matches 137; Conservative 68; Mismatches 144; Indels 52; Gaps 9;  
QY 2 RHSHYFQWLSPLLSVAVTQOLYAQNESLPTVELEPVVITIDKSGMALANRITQMPHT 60  
Db 94 RDEHTF--IVVPASEAAVPAQARSEP-----LDMQMEITASRTSSDLVSATRQ----- 141  
QY 61 TKVIEEIQEQATGSRQLADVMAQLIPSLGVSSTTSNFGQTMHGRQVQFLNGVPLTG 120  
Db 142 STVIEHAQLEELROGSDSLATVAKAVPGMSDSRTITEYQTLGRSMLVMVGVPLNT 201  
QY 121 SRDISRQLNSINPNQVARIIVLSGATSIYGSATGGLINIVTKSDLEEEQETRIGVHGS 179  
Db 202 NRDSRNLANDPALIERIEVIRGSAIYGSATGGIISITRPPAGENRAETRLSATSP 261  
QY 180 -SKLSSEGIGVQVGSVAGVSENGVNLARDVDY-----RTTGGAFDANGKRIAPPAQTD 234  
Db 262 LTRLSGDLGGGQFOYFAG-----SLGALDYSFDFGRHVGSYDAHDRIAPPEPSQGD 315  
QY 235 QDSKSLSVNTNVDMQLDDKQINLALTHYNDKQDTPDYAPDYGNNRLAVLFGKPSLNAIK 294  
Db 316 LFDENVNIGKGLGRIDENQVQLASHDYARDQTDYATD--PRVRLPGSPVPANAIK 373  
QY 295 GLSLSEQPKTKSTFNINHHDDLWGNTINTNAYRREKGRFPFVAPFSTAKALPILQSN 354

Db	374	GLEDEQNRIRNTLANLEVENLIDILSRSLVAQLYTYDDYTRTPFPDA-----	420
Qy	355	MNLPSATLDAVYTKAPQARAYGVLOSKEKSLVIGRVENLNKP	395
Db	421	-----RAVSTRGNGVDQIMQNSBVFESRLTLTRTP	449
RESULT 3			
AAV72914	ID	AAV72914 standard; Protein; 725 AA.	
AAV72914;			
13-JUN-2001	(first entry)		
E. coli iroNec	extracytoplasmic protein fragment.		
ironEc;	extracytoplasmic protein; immunogen; vaccine; E1; UTI;		
immunotherapy; extraintestinal infection; urinary tract infection;			
meningitis; pneumonia; intra-abdominal infection; antibiotic.			
Escherichia coli.			
Key	Location/Qualifiers		
Peptide	1..24		
Protein	/label= Signal_peptide		
	25..725		
Misc-difference 120..121	/label= Mature_E. coli_iroNec_extracytoplasmic_protein		
	/note= "Encoded by GTG CGT TAT"		
Misc-difference 134..136	/note= "Encoded by AAC TGG"		
WO200121636-A1.			
29-MAR-2001.			
22-SEP-2000; 2000WO-US26117.			
22-SEP-1999; 99US-0155621.			
(UYNY ) UNIV NEW YORK STATE RES FOUND.			
Russo T, Carlino U;			
WPI; 2001-244936/25.			
N-PSDB; AAD03054.			
Novel isolated iroNec polynucleotide from extraintestinal isolate of			
Escherichia coli useful as vaccine for treating or preventing			
extraintestinal infections caused by extraintestinal pathogenic			
Escherichia coli -			
Claim 9; Page 39-41; 44pp; English.			
The invention relates to the identification of iroNec gene,			
from an extraintestinal isolate of Escherichia coli. This			
gene is expressed in increased amounts in human urine and is			
identified by transposon (TnpA) mutagenesis. iroNec gene			
encodes an extracytoplasmic protein. This gene can be used as			
an immunogen in vaccine formulations. The recombinant vector			
comprising nucleotide sequence encoding one or more antigenic			
epitope of iroNec is useful for diagnostic and immunotherapeutic			
purposes. The iroNec antigenic peptide is useful for treating or			
preventing extraintestinal infections (EIs) caused by extra-			
intestinal pathogenic E. coli (ExPEC). The EIs include urinary			
tract infection (UTI), meningitis, intra-abdominal infection			
and pneumonia.			
The present sequence is E. coli iroNec extracytoplasmic protein.			
Sequence 725 AA;			

	Query Match	9.4%: Score 191.5; DB 22; Length 725;
	Best Local Similarity 24.6%; Pred. No. 3.6e-07;	
	Matches 84; Conservative 59; Mismatches 135; Indels 63; Gaps 17;	
Qy	9 WLSPLPLSVATTOOL---VAQPNESLPTVELPEVITIDKSGMALANRITOMPHTTKVI 64	
Db	8 W-SLTVLLVLVGLNSQVSAVKYEDDDND-----ETLVVE-----ATAQVLKQPGSVVI 54	
Qy	65 YEEOIQEATGSRQLADYMAQLIPSLGV---SSGTTSNPQGT---MHGRVOEFLNGV 116	
Db	55 TSEDI-KKTPPVNDISDITIRHW-PGVNLTGNSASGTRGNNNRQIDIRGMGPENTILLIGV 112	
Qy	117 PLT-----GSRDISRQNS--INPQVARIETLGS-ARSIYSGATGGLINIVTSD 165	
Db	113 PVTSRNSVYSNRGEEDTRGDTNRWVPEQVERIEIRGPAARQVSGAAGVNIITKRP 172	
Qy	166 LEEQFETRIGVHSGSKLSESGIGYOVQGVSGVSENGVLAFLDVIDY-RTTGAFDANGK 224	
Db	173 TNDWHTGSLSTLTNPQESSEEGATRRANSLSGPLAGDALITRLVGNLAKTIDADSDIN-- 230	
Qy	225 RIAPPAQTDK---QDSKLSLVNTVDWQLDKQINILALTHYNDKQDTPYADPYGNRL 280	
Db	231 ----SPVGTKNAGHGVRNKGINDGVSMKLNPOQILDFEYGY--SRQGNIVAGDTQSS 284	
Qy	261 AVLREKPSLNAIKGLSLSEQPKTKSFNNY--HHDDWM 319	
Db	265 SSAVTE-----SLAKSGKETNRLRYONGITTHGIW 315	
RESULT 4		
AA134501		
ID	AA134501 standard; Protein; 708 AA.	
XX	AA134501;	
XX	25-AUG-1999 (first entry)	
XX	25-AUG-1999 (first entry)	
XX	Porphyromonas gingivalis protein PG40.	
XX	Porphyromonas gingivalis; PG, periodontal disease; gingivitis;	
XX	Porphyromonas gingivalis.	
XX	Porphyromonas gingivalis.	
XX	W09923870-A1.	
XX	17-JUN-1999.	
XX	10-DEC-1998; 98W0-AU01023.	
XX	04-AUG-1998; 98AU-0005028.	
XX	10-DEC-1997; 97AU-0000839.	
XX	31-DEC-1997; 97AU-0001182.	
XX	30-JAN-1998; 98AU-0001546.	
XX	10-MAR-1998; 98AU-0002264.	
XX	09-APR-1998; 98AU-0003128.	
XX	23-APR-1998; 98AU-0003128.	
XX	05-MAY-1998; 98AU-0003336.	
XX	22-MAY-1998; 98AU-0003654.	
XX	29-JUN-1998; 98AU-0004917.	
XX	(CSLC-) CSL LTD.	
XX	Aglus CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;	
XX	Ross BC, Rochel LJ, Webb EA;	
XX	WPI; 1999-385613/32.	
XX	N-PSDB; AAX91719.	
XX	Antigenic Porphyromonas gingivalis peptides for preventing	
XX	gingivitis	
XX	Claim 1; Page 482-483; 586pp; English.	





KW Porphyromonas gingivalis; PG, periodontal disease; gingivitis;  
KW vaccine; antigenic.  
OS Porphyromonas gingivalis.  
PN WO929870-A1.  
PD 17-JUN-1999.  
PP 10-DEC-1998; 98WO-AU01023.  
PR 04-AUG-1998; 98AU-0005028.  
PR 10-DEC-1997; 97AU-0000839.  
PR 31-DEC-1997; 97AU-0001182.  
PR 30-JAN-1998; 98AU-0001546.  
PR 10-MAR-1998; 98AU-0002264.  
PR 09-APR-1998; 98AU-0002911.  
PR 23-APR-1998; 98AU-0003128.  
PR 05-MAY-1998; 98AU-0003338.  
PR 22-MAY-1998; 98AU-0003654.  
PR 29-JUL-1998; 98AU-0004917.  
PX (CSLC-) CSL LTD.  
PA Aglus CT, Barr IG, Hocking DM, Margetts MB, Patterson MA,  
PI Rose BC, Rothel LJ, Webb EA;  
PL WP1; 1999-385613/32.  
DR N-PSDB; AAX91690.  
XX Antigenic Porphyromonas gingivalis peptides for preventing  
XX gingivitis  
XX Claim 1; Page 450-451; 588pp; English.  
XX  
XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic  
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAX914318 to  
CC AAX914583. AAX91802 to AAX91989 represent PCR primers used in the  
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial  
CC activity with a vaccine mechanism of action. The PG polypeptides can be  
CC used as vaccines especially against Porphyromonas gingivalis. Probes can  
CC be used to detect Porphyromonas gingivalis in standard hybridisation  
CC assays. Porphyromonas gingivalis is involved in periodontal disease  
CC especially gingivitis.  
CC  
CC Sequence 757 AA:  
XX SQ

Query Match 8.7%; Score 178.5; DB 20; Length 757;  
Best Local Similarity 25.1%; Pred. No. 4,6e-06;  
Matches 83; Conservative 56; Mismatches 147; Indels 45; Gaps 16.

QY 31 LEFTELPEPVITIDKSGMLANRITOMPHTKIYEEOIQEATGSRDLADVAO-LIPS 89  
Dbb | | | | | : | : | : | : | : | : | : | : | : | : | : |  
106 LRTNNLEEVVTV-----GTGRTRLVDAVPATEVLTKOI---ASFAPTSALLGLSPS 158  
QY 90 LGVSSGTTSNFGQT--MHGRVOFLNGVPLTGSRDISRLNSINNOVARIEVLSGA-T 146  
Dbb | | | | | : | : | : | : | : | : | : | : | : | : | : |  
Ddb 159 FDFGFPMWSFMQLNGLSKRYILLHDGRKYVDVGQADLSRI SPDQRLELVKGASS 218  
QY 147 STYGSGATGGLINIYTKSLDEEQFT--RIGVHSGKLSEGGYGVGVSAGVENGNV 204  
| | | | | : | : | : | : | : | : | : | : | : | : | : |  
Dbb 219 STYSGDALAGVINIVTYKNTNRISA VTSHRISKYNDROTNTSIDNIIGK----PSSNTNV 274  
QY 205 LARLDVDVYTTG---GAFPDANGRIAPSPA--QTDKODSKLSLVNTVMOLDDKNIMV 259  
Dbb 275 FF-----YHTDMQNSPFIIKKKKSGGEVLETTYKTTRRA-QENOGVQSLSYATNMV 328  
QY 260 ALT---HYNDKODPTDYAPDYGNRLAVLFGEKPSLNAIKGL-----SLSEQPTTKS 307  
Ddb 329 SFSGNVQVKKRO--IFTTFSEKKAYDMOYR-ALTASLTNYLFPNGHLTLSDAVYDRF 385  
QY 308 TFININHDDLMGNTINTNTAAYRRREKRGFRYP 338  
| | | | | : | : | : | : | : | : | : | : | : | : | : |

Dd		386	RFGVLYHDKDSESLINNQG--QQEQPIFFP	414
<b>RESULT 7</b>				
ID	AAV34348	AAV34348	standard; Protein; 763 AA.	
XX	AAV34348;			
AC				
DT	25-AUG-1999	(first entry)		
XX				
DE	Porphorymonas gingivalis protein PG13.			
XX				
KW	Porphorymonas gingivalis; PG; periodontal disease; gingivitis; vaccine; antigenic.			
OS	Porphorymonas gingivalis.			
XX				
PN	WO9929870-A1.			
PD	17-JUN-1999.			
PF	10-DEC-1998;	98WO-AU01023.		
XX				
PR	04-AUG-1998;	98AU-0005028.		
PR	10-DEC-1997;	97AU-0000839.		
PR	31-DEC-1997;	97AU-0001182.		
PR	30-JAN-1998;	98AU-0001546.		
PR	10-MAR-1998;	98AU-0002264.		
PR	09-APR-1998;	98AU-0002911.		
PR	23-APR-1998;	98AU-0003128.		
PR	05-MAY-1998;	98AU-0003338.		
PR	22-MAY-1998;	98AU-0003654.		
PR	29-JUL-1998;	98AU-0004917.		
XX				
PA	(CSLC-) CSL LTD.			
PT	Agius CT, Barr IG, Hocking DM, Margeltes MB, Patterson MA;			
PI	Ross BC, Roedel LJ, Webb EA;			
XX				
DR	MFI; 1999-385613/32.			
N-PADB; AAX91566.				
XX				
PT	Antigenic Porphorymonas gingivalis peptides for preventing gingivitis			
PS	Claim 1; Page 308-309; 588pp; English.			
CC	AAV91536 to AAV91801 encode two hundred and sixty six antigenic Porphyromonas gingivalis (PG) polypeptide sequences given in AAV34318 to AAV34563. AAV91802 to AAV91989 represent PCR primers used in the isolation of the PG polypeptides. The PG polypeptides have antibacterial activity with a vaccine mechanism of action. The PG polypeptides can be used as vaccines especially against Porphyromonas gingivalis. Probes can be used to detect Porphyromonas gingivalis in standard hybridisation assays. Porphyromonas gingivalis is involved in periodontal disease especially gingivitis.			
SQ	Sequence 763 AA;			
Query Match	8.7%; Score 178.5; DB 20; Length 763;			
Best local similarity	25.1%; Pred.No. 4.6e-06;			
Matches	83; Conservativity 56; Mismatches 147; Indels 45; Gaps 16;			
Oy	31 LFTVALPEPVITIDSGMAIANRIQTOMHTTKYVEEDIQEATGSRQLADVMAG-LIPS	89		
Dd	112 LFTNNLEEVVT---GGTTRYRLVDAPAIVAEVLAKDI--ASFSAPTSEXLLGLSPS	164		
Oy	90 LGVSGETTSNFGOT--MHGRGVDFLNGCVPLTGSRDISROLNSINPNQVARIEVLSGA-T	146		
Dd	165 PDGFNNMGSSFRQLNLGLSKYLILIIDKKRYVGDPGGAGADSRSIPDDIERELVKASS	224		
Oy	147 STYGSCATGGLINITKSDLEEDEFET--RGVHGSKLSSBECIGYOVCQSVAQVSENQNV	204		

Db 225 SLYGDAIAGVINTKNTNLSATSHRISKYNDRTQNTSLDINIK-----FSNTNY 280  
Qy 205 LARLDVYRTTG---GAFDANGKRIAPEPA--QTDKDSKSLSVNTNVDWQDDKONINL 259  
Db 281 FF-----YHTDGMQNSPFEIKKKKSGEPVLEETKYKTFRA-QENQGVQSLSYYATNNL 334  
Qy 260 ALT-----HYNDKQDTYADYGNRLAVLFGEKPSLNAIKGL-----SLSEQKTKYS 307  
Db 335 SFGSGNVQYKRO--IFTTFSEKKAYDMYR-ALTASLTGNTLYLFFNGLHTLSFDVAYDRF 391  
Qy 308 TFINVYHDDLWGNTINTNAYVYRREKGRFP 338  
Db 392 RFGYLYHDKSSESLINNOG---QTEQPTFFP 420  
RESULT 8  
ABG17860  
ID ABG17860 standard; Protein; 454 AA.  
XX AC ABG17860;  
XX DT 18-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #17851.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX PN WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US08631.  
XX PR 31-MAR-2000; 2000US-0540217.  
XX PR 23-AUG-2000; 2000US-0649167.  
XX PA (HYSEQ-) HYSEQ INC.  
XX PI Dmanac RT, Liu C, Tang YT;  
XX DR WPI; 2001-639362/73.  
XX DR N-PSDB; AAS82047.  
XX PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX PS Claim 20; SEQ ID NO 48219; 103pp; English.  
XX CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (II) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 454 AA;  
Query Match 8.4%; Score 171.5; DB 22; Length 454;  
Best Local Similarity 22.8%; Pred. No. 8.5e-06;  
Matches 92; Conservative 49; Mismatches 141; Indels 121; Gaps 19;  
Qy 17 VAVTQQLYAQNESLPTVELEP-----VVITIDKSGMALANRITQMHTTKVI 64  
Db 66 MSVYAQAAVEPKEDTITVTAAPAQESAWGPAATIAARQSATGTCTDTPIQKVQSISV 125  
Qy 65 YEEIQEQATGSRQLADVMAQLIFSLGVSSGT--TSNP-----GOTMHRQVQFLLNG 115  
Db 126 TAEEM-----ALHOPKSVKEALSVTPGVSVGTRGASNTYDHLIIRGFAAQSQNNYLLNG 180  
Qy 116 VPLTGS--RDISRQLNSINPNQVARIIEVLGATSI-YGSGATGGLINIVTK-----SDLEE 168  
Db 181 LKQGNFYND-----VIDPYMLERAETMRGPVSVLYGKSNPGGLNNMVKRPTTEPLKE 235  
Qy 169 EQFETRIGVHSGSKLSSEGGYGVQSVAGVSENGVNLARLDVYRTTGGAFDANGKRIAP 228  
Db 236 VQP-----KAGTDSL-FQTGDFSDSLDDGVYS-----YRLTGLARSAN----- 274  
Qy 229 EPAQTDKQDSKSLSVNTNVDWQDDKONINLALTHYNDKQDTYADYGNRLAVLFGEKP 288  
Db 275 --AQQKSGEQRYAIAPAFTWRPDDKTNFTF-LSVFQNEPDTGY-----YGWLP 320  
Qy 289 SLNAIKGLSLSEQPKTKSTFNINYHDDLLWGNTINTNAYVYRREKGRFPYFPVAPPSIAKA 348  
Db 321 KEGTVEPL---PNGKRLPTDFN-----DRAKNAYSNRNEK----- 352  
Qy 349 LPILQSNLPSATLDAYTKAPARAYGVLOSKEAELGRVFN 391  
Db 353 -----MN-----EPETGYGWLPKDGTVEPL---PN 375  
RESULT 9  
ABBS2669  
ID ABBS2669 standard; Protein; 660 AA.  
XX AC ABBS2669;  
XX DT 11-FEB-2002 (first entry)  
XX DE Escherichia coli polypeptide SEQ ID NO 732.  
XX KW Escherichia coli; B2/D-A; antiinflammatory; antibacterial;  
KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;  
KW systemic infection; non-diarrhoeal infection; septicaemia;  
KW pyelonephritis; antibiotic resistance.  
XX OS Escherichia coli.  
XX PN WO200165572-A2.  
XX PD 13-SEP-2001.  
XX PF 12-MAR-2001; 2001WO-EP03445.  
XX PR 10-MAR-2000; 2000FR-0003145.  
XX PR 02-FEB-2001; 2001FR-0001449.  
XX PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
XX PI Bingen E, Bonacora S, Clermont O, Nassif X, Tinsley C;  
XX DR WPI; 2001-550253/61.  
XX PT A library of DNA fragments of Escherichia coli strains for the  
PT phylogenetic determination of a given strain comprises polynucleotides of  
PT nature B2/D+ A- -

XX Example 6, Fig 6, 646pp, English.  
XX  
XX The invention relates to a library of DNA fragments of *Escherichia coli*  
CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA889533)  
CC and encoded proteins (AB852459-AB852919 and AB852954-AB853094) of nature  
CC B2/D-A. The polynucleotides have potential anti-inflammatory,  
CC antibacterial and immunosuppressive activity as part of pharmaceutical  
CC compositions used to treat, palliate or prevent extra-intestinal *E. coli*  
CC infections. The polypeptides are useful for determining the phylogenetic  
CC group of a given *E. coli* strain. These polypeptides can detect and treat  
CC an undesired development of *E. coli*, particularly an extra-intestinal  
CC infection that include systemic and non-diatracheal infections such as  
CC septicaemia, pyelonephritis and meningitis this is particularly  
CC advantageous as bacterial resistance is increasing with the more  
CC frequent use of broad spectrum antibiotics.  
XX  
XX  
SQ Sequence 660 AA:  
Query Match 7.3%; Score 149; DB 22; Length 660;  
Best Local Similarity 26.7%; Pred. No. 0.0011;  
Matches 74; Conservative 48; Mismatches 103; Indels 52; Gaps 16;  
QY 7 FQMLSPILSVATQQLYAPNPSLPVLEPVVITIDKSGMALANRITQMPHTTKYIE 66  
DB 6 FTSLRLSLALAV-----SATLPFPATETMTVTATGNARSS--FEAPMVSVITDT 55  
QY 67 EQIOEATGSRQADVAQAOLIPSLGV--SSGTSNFGQTMHG---ROYQPLNGVPL---T 119  
DB 56 SAENQATAT--ATDLIRIVPGITLDTGRITNGQDVNMGYDHRYVLVVDGRCQDTDT 112  
QY 120 GSRDISRQINS--INPNQVARIIVLSGATS--YSGGATGGL--NIVTKSDELEBPETR 174  
DB 113 G-----HNGTFLDPLALIKRVEIVRGPALVYSGGALGVISYDTVDADLLQEGSSG 166  
QY 175 IGVHGS-KLSSEGIYGVGQSVAGVSENGVLAFLD---VDYRTTGAGDANGKRIAPR 229  
DB 167 FRFFGTGTGDHSLG--LGASAFGRTE-----LDGIIVAMSSRDRLRQNSG-----E 213  
QY 230 PAQTDKODSKLSVNTVNDWQLDKONINLALTHYND 266  
DB 214 TAPNDESINMLAKGT---WQIDSAQSLSGIVRYNN 247  
RESULT 10  
ABG17918  
ID ABG17918 standard; Protein; 718 AA.  
AC ABG17918;  
XX  
XX 18-FEB-2002 (first entry)  
XX  
XX Novel human diagnostic protein #17909.  
XX  
XX Human; Chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
OS  
PN MO200175067-A2.  
XX  
XX 11-OCT-2001.  
XX  
XX 30-MAR-2001; 2001MO-US08631.  
XX  
XX 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSB-) HYSEQ INC.  
XX  
XX Drmanac RT, Liu C, Tang YT;  
XX  
XX WFI, 2001-639362/73.  
DR

DR N-PSDB; AAS82105.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostic, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 20; SEQ ID NO 48277; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridization probes,  
CC polymerase chain reaction (PCR) primers, oligomers and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIFO  
CC at ftp.wifo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 718 AA:  
Query Match 7.2%; Score 147.5; DB 22; Length 718;  
Best Local Similarity 21.5%; Pred. No. 0.0016;  
Matches 92; Conservative 68; Mismatches 152; Indels 115; Gaps 21;  
QY 26 QPN--ESLPTVLEPVVITID-----KSGMALANRITQMPHTT 61  
DB 176 KRGVYSHTVLVGPQKTVBEIQRTQKHEDALGVAAHWKYEBAAG--ARSGHD 233  
QY 62 KYIEBQI---OEATGSRQADVAQAOLIPSLGVSSGTSNFGQ-----TMHG 106  
DB 234 RIAMLRKLIAMQEMADSGEMLDEVRSQVL-----TTGCTSYANQPRITPLIRGFPAEG 288  
QY 107 RQVQPLNGVPLTGSRDISRLNSINPNQVARIIVLSGATS--YSGGATGGLINITYK-- 163  
DB 289 QSQNNYTLNGKL--ODNFYNDVAVIDPYMLERAEIIRGVSVLYGKSSFGGLNWSKRP 345  
QY 164 --SDLEEQFETRIGVHGSKLSSEGIYGVGQSVAGVSENGVLAFLD--VDYRTTGAGAFDA 221  
DB 346 TTEPLKEVQF-----KAGTDSL--FQRTGDFPSGLDDDDGVYS-----YRLTGLARSA 390  
QY 222 NGRIAPBPAQTDKODSKLSVNTVNDWQLDKONINLALTHYNDKQTDYAPDYGNRLA 281  
DB 391 N-----AQCGSEBQGVYAIAPFTMRPDDXTNFTF--LSYFQNPENRGY----- 432  
QY 282 VLFGEKPSLNAIKGLSLSPQPKTKSTFNINHHDDIMKNTINTNAYYRREKGRFY--PFV 340  
DB 433 --YGMUPKEGTVBPLPNAQDLTS-----GRVISGSGTLLTMQVARIIDPHR 476  
QY 341 APF-----SIKALPTIQSNNLPSATLDAYTKAPQARAVGVQSESKAE--VLGRVP--NL 392  
DB 477 KTFGSGKIRQLMRLQLEMLSKREIKTYLVTAPLA---GTLOGIGAAWYVYIGKSPANL 533  
QY 393 NKPKRAL 399  
DB 534 SYSEAM 540  
RESULT 11  
AAR42451  
ID AAR42451 standard; Protein; 970 AA.  
DR

XX AC AAR42451;  
 XX DT 27-MAY-1994 (first entry)  
 XX DE Enzyme involved in eicosapentaenoic acid (EPA) synthesis.  
 XX KW EPA; eicosapentaenoic acid synthetase; drug; anticoagulant;  
 KW hypolipemic; hypoglycemic; antihypertensive; anticancer; pesticide;  
 KW foodstuff; additive.  
 XX OS Shewanella putrefaciens.  
 XX PN WO9323545-A.  
 XX PD 25-NOV-1993.  
 XX PF 14-MAY-1993; 93WO-JP00641.  
 XX PR 15-MAY-1992; 92JP-0147945.  
 XX PA (SAGA ) SAGAMI CHEM RES CENTRE.  
 XX PI Kato S, Kondo K, Yamada A, Yazawa K;  
 XX WI; 1993-386577/48.  
 XX DR N-PSDB; AAR51128.  
 XX PT Gene coding for eicosa-penta:enoic acid synthetase - is isolated  
 PT from Pseudomonas, Alteromonas or Shewanella and used for  
 PT recombinant prodn. of eicosa-penta:enoic acid  
 XX Claim 5; Page 45-51; 106pp; Japanese.  
 XX PS EPA is useful as a drug, having anticoagulant, hypolipemic,  
 CC hypoglycemic, antihypertensive and anticancer activity. It is also  
 CC a pesticide and is useful as a nutritional foodstuff and animal feed  
 CC additive.  
 XX SQ Sequence 970 AA;  
 Query Match 7.2%; Score 147.5; DB 14; Length 970;  
 Best Local Similarity 23.3%; Pred. No. 0.0024;  
 Matches 53; Conservative 51; Mismatches 72; Indels 51; Gaps 10;  
 QY 15 LSVAVTQOLYAQPNEISLPTVELEPVVITIDKSGMALANRITQMPHTTKVIYEEQIQEQAT 74  
 DB 21 LTASLAMPVFAE--ETAEEQIERVAVT-----GSRIAKAELTQAPVVSLSAEELTK--F 72  
 QY 75 GSRQLADVNAQLIPSLGV-----SSGTTSNFGQTMHGRQVQFLNG----VPL 118  
 DB 73 GNQDLGSLVLAEL-PAIGATNTIIGNNNSNSAGVSSADLRRLGANRTLVLVNGKRYVAGQ 131  
 QY 119 TGSRDISRQLNSINPNQVARIIVLS-GATSIYCGATGGLINIVTKSLDEEBEQPETRIGV 177  
 DB 132 PGSAEV--DLSTIPTSMISRVEIVTGGASAIYGSADVSGVINVLKEDFEGFEFNAR--- 186  
 QY 178 HGSKLSSEGIGVQGVQSVAGVSENGVNLARLDVYRTTGGAFDANGK 224  
 DB 187 --TSGSTESVGTQ-----EHSFDILGGANVADGR 213  
 RESULT 12  
 AAR99460  
 ID AAR99460 standard; Protein; 970 AA.  
 XX AAR99460;  
 XX DT 30-JAN-1997 (first entry)  
 XX DE Biosynthetic enzyme of icosapentaenoic acid synthase.  
 XX KW Icosapentaenoic acid synthase; EPA; drugs; agrochemicals;

KW foodstuffs; animal feed; lipid balance correction; antihypertensive;  
 KW antiinflammatory; anticancer agent.  
 XX OS Shewanella putrefaciens.  
 XX PN WO9621735-A1.  
 XX PD 18-JUL-1996.  
 XX PF 12-JAN-1996; 96WO-JP00030.  
 XX PR 13-JAN-1995; 95JP-0004299.  
 XX PA (SAGA ) SAGAMI CHEM RES CENTRE.  
 XX PI Kato S, Kondo K, Yamada A, Yazawa K;  
 XX WI; 1996-342288/34.  
 XX DR N-PSDB; AAT34137.  
 XX PT Production of icosapentaenoic acid using transformed E. coli - uses  
 PT DNA coding for icosapentaenoic acid synthase derived from Shewanella  
 PT strain  
 XX Claim 5; Page 57-61; 145pp; English.  
 XX CC The DNA sequence (AAT34137) which encodes the biosynthetic enzymes of  
 CC icosapentaenoic acid (EPA) can be used to transform Escherichia coli.  
 CC The DNA sequence allows efficient microbial production of EPA, which  
 CC is a raw material for drugs, agrochemicals, foods and animal  
 CC feedstuffs. EPA is also useful for lipid balance correction and as  
 CC an antihypertensive, antiinflammatory and anticancer agent.  
 XX SQ Sequence 970 AA;  
 Query Match 7.2%; Score 147.5; DB 17; Length 970;  
 Best Local Similarity 23.3%; Pred. No. 0.0024;  
 Matches 53; Conservative 51; Mismatches 72; Indels 51; Gaps 10;  
 QY 15 LSVAVTQOLYAQPNEISLPTVELEPVVITIDKSGMALANRITQMPHTTKVIYEEQIQEQAT 74  
 DB 21 LTASLAMPVFAE--ETAEEQIERVAVT-----GSRIAKAELTQAPVVSLSAEELTK--F 72  
 QY 75 GSRQLADVNAQLIPSLGV-----SSGTTSNFGQTMHGRQVQFLNG----VPL 118  
 DB 73 GNQDLGSLVLAEL-PAIGATNTIIGNNNSNSAGVSSADLRRLGANRTLVLVNGKRYVAGQ 131  
 QY 119 TGSRDISRQLNSINPNQVARIIVLS-GATSIYCGATGGLINIVTKSLDEEBEQPETRIGV 177  
 DB 132 PGSAEV--DLSTIPTSMISRVEIVTGGASAIYGSADVSGVINVLKEDFEGFEFNAR--- 186  
 QY 178 HGSKLSSEGIGVQGVQSVAGVSENGVNLARLDVYRTTGGAFDANGK 224  
 DB 187 --TSGSTESVGTQ-----EHSFDILGGANVADGR 213  
 RESULT 13  
 AAW37048  
 ID AAW37048 standard; Protein; 970 AA.  
 XX AAW37048;  
 XX DT 03-JUL-1998 (first entry)  
 XX DE S. putrefaciens EPO biosynthesis gene cluster ORF4 product.  
 XX KW SCRC-2874; FERM BP-1625; eicosapentaenoic acid; EPA;  
 KW biosynthesis gene cluster; synthetase.  
 XX OS Shewanella putrefaciens.  
 XX PN WO9801565-A1.

PD 15-JAN-1998.  
 XX  
 XX 09-JUL-1997; 97WO-JP02371.  
 XX  
 XX 10-JUL-1996; 96JP-0180845.  
 PR  
 XX (SAGA ) SAGAMI CHEM RES CENTRE.  
 PA  
 XX Kato S, Kondo K, Yamada A, Yazawa K;  
 PI  
 XX MPI: 1998-101060/09.  
 DR N-PSDB; AAV00503.  
 XX  
 XX Bicosapentaenoic acid produced by culture of transformed *Escherichia*  
 PT coli - containing an eicosapentaenoic acid synthetase gene derived  
 PT from the marine microorganism *Shewanella*  
 XX  
 XX Example 1; Pages 44-51; 110pp; Japanese.  
 PS  
 XX The present sequence is encoded by the *Shewanella putrefaciens*  
 CC SCRC-2874 (FERM BP-1625) eicosapentaenoic acid (EPA) biosynthesis  
 CC gene cluster.  
 CC A novel EPA (useful in drugs, pesticides, foods and feedstuffs) is  
 CC encoded by synthetase enzyme gene sequences comprising parts of the  
 CC full sequence of the synthetase gene from the marine microorganism  
 CC *S. putrefaciens* SCRC-2874 (FERM BP-1625), in which at least 1 of  
 CC the 9 open reading frames (ORF) (numbered 2-10) in the gene have  
 CC been deleted. In particular the gene sequences comprising the  
 CC following parts of the full gene:  
 CC (1) bases 8081-9441, 12314-13084 and 13889-32520;  
 CC (2) bases 8081-9441, 12314-13084, 13889-32520 and 34627-35559;  
 CC (3) bases 8081-9441, 12314-13084 and 13889-35559;  
 CC (4) bases 8081-9441, 9681-13084 and 13889-32520;  
 CC (5) bases 8081-9441, 9681-13084, 13889-32520 and 34627-35564; and  
 CC (6) bases 8081-9441, 9681-13084 and 13889-35564, are claimed.  
 CC  
 XX Sequence 970 AA;  
 SQ  
 Query Match 7.2%; Score 147.5; DB 19; Length 970;  
 Best Local Similarity 23.3%; Pred. No. 0.0024;  
 Matches 53; Conservative 51; Mismatches 72; Indels 51; Gaps 10;  
 QY 15 LSAVAATQQLYAQNESLPTVELEPVVITIDKSGMALANRITOMPHTKVYERQIOEQAT 74  
 DB 21 LTAALMPVAF--ETAAEQIERVAVT---GSRJAKAEITQAPVVISAEELTK--F 72  
 QY 75 GSRQADVMAQLIPSLGV-----SSGTTNFGQTMHGRQVOFLNG---VPL 118  
 DB 73 GNDDLSGLAEL-PAIGATNTTIIGNNSSNSAGVSSADRLRLGANRTLVVNGRRYVAGQ 131  
 QY 119 TGSRDISRQNSINPQVARIIVLS-GATSIYSGATGGLINTVTKSDLEBOFETRIGV 177  
 DB 132 PGSAEV--DLSTPTSMISRVEIVTGASAIYGSDAVSGVINVLKEDFEGEFENAR--- 186  
 QY 178 HGSRLSSEGIQYGVGSGVAGSENGVNLARDVDYRTTGAFDANGK 224  
 DB 187 --TSGSTESVGTQ-----EHSFDILGANVADGR 213  
 RESULT 14  
 AAM89398  
 ID AAM89398 standard; Protein; 970 AA.  
 XX  
 XX AAM89398;  
 AC  
 XX 19-OCT-1999 (first entry)  
 DT  
 XX  
 XX S. putrefaciens PKS-like cluster ORF 4 protein.  
 DE  
 XX Polyketide-like synthesis; PKS; PKS-like gene; PUFA; DHA; transgenic;  
 KM poly-unsaturated fatty acid; eicosapentaenoic acid; docosahexaenoic acid;  
 KM EPA; oil; dietary supplement; infant feeding formulation; malnutrition;  
 KM intravenous feeding formulation; cooking oil; fat; anti-inflammatory;

KW cholesterol; open reading frame; ORF.  
 XX  
 XX *Shewanella putrefaciens*.  
 OS  
 XX WO9855625-A1.  
 PN  
 XX 10-DEC-1998.  
 PD  
 XX 04-JUN-1998; 98WO-US11639.  
 PF  
 XX 04-JUN-1997; 97US-0048650.  
 PR  
 XX (CALJ ) CALGENE LLC.  
 PA  
 XX Faciottti D, Laesner M, Metz JG;  
 PI  
 XX MPI: 1998-070271/06.  
 DR N-PSDB; AAV81945.  
 XX  
 XX New nucleic acid encoding polyketide-like synthesis enzymes of  
 PT *Vibrio marinus* - and transformed plants and microbes that produce  
 PT polyunsaturated fatty acids, useful as pharmaceuticals and food  
 PT supplements  
 XX  
 XX Example 1; Fig 4; 153pp; English.  
 PS  
 XX The invention provides polyketide-like synthesis (PKS)-like genes that  
 CC are used for the production of long chain poly-unsaturated fatty acid  
 CC (PUFA) productions. Genes responsible for eicosapentaenoic acid (EPA)  
 CC production in *Shewanella putrefaciens* and novel genes associated with  
 CC the production of docosahexaenoic acid (DHA) in *Vibrio marinus* are used  
 CC to generate transgenic plants that can express transgenes encoding  
 CC PKS-like genes associated with PUFA production. The PKS-like genes are  
 CC used to transform plants and microbial cells to give recombinants having  
 CC altered contents of PUFA (specifically DHA and EPA). Oils from these  
 CC plants are useful as dietary supplements (in infant feeding  
 CC formulations, to give a PUFA profile closer to that of human milk; for  
 CC treating malnutrition; in intravenous feeding formulations; in cooking  
 CC oils, fats etc.); also as anti-inflammatory agents and for reducing  
 CC cholesterol levels. Fragments from the genes are useful as probes to  
 CC isolate related molecules or to detect organisms that express PKS-like  
 CC genes. The method facilitates large scale production of PUFA by providing  
 CC new pathways for their synthesis or suppressing interfering pathways.  
 CC Expression of PUFA in seeds allows simple recovery, as oil which can be  
 CC engineered to have a particular PUFA profile. Expression in microbes also  
 CC allows simple recovery and control of PUFA profile and is not subject to  
 CC external variables such as weather or food supply. Sequences  
 CC AAM89396-404 represent different ORF proteins of *S. putrefaciens* PKS-like  
 CC cluster.  
 CC  
 XX Sequence 970 AA;  
 SQ  
 Query Match 7.2%; Score 147.5; DB 20; Length 970;  
 Best Local Similarity 23.3%; Pred. No. 0.0024;  
 Matches 53; Conservative 51; Mismatches 72; Indels 51; Gaps 10;  
 QY 15 LSAVAATQQLYAQNESLPTVELEPVVITIDKSGMALANRITOMPHTKVYERQIOEQAT 74  
 DB 21 LTAALMPVAF--ETAAEQIERVAVT---GSRJAKAEITQAPVVISAEELTK--F 72  
 QY 75 GSRQADVMAQLIPSLGV-----SSGTTNFGQTMHGRQVOFLNG---VPL 118  
 DB 73 GNDDLSGLAEL-PAIGATNTTIIGNNSSNSAGVSSADRLRLGANRTLVVNGRRYVAGQ 131  
 QY 119 TGSRDISRQNSINPQVARIIVLS-GATSIYSGATGGLINTVTKSDLEBOFETRIGV 177  
 DB 132 PGSAEV--DLSTPTSMISRVEIVTGASAIYGSDAVSGVINVLKEDFEGEFENAR--- 186  
 QY 178 HGSRLSSEGIQYGVGSGVAGSENGVNLARDVDYRTTGAFDANGK 224  
 DB 187 --TSGSTESVGTQ-----EHSFDILGANVADGR 213

RESULT 15  
AAB10465  
ID AAB10465 standard; Protein; 970 AA.  
XX  
AC AAB10465;  
XX  
DT 11-DEC-2000 (first entry)  
XX  
DE Shewanella putrefaciens PKS protein ORF4.  
XX  
KW PKS pathway; polyunsaturated long chain fatty acid; plant; transgenic;  
KW polyketide-like synthesis; PUF; dietary supplement; intravenous feeding;  
KW malnutrition; cooking oil; cooking fat; margarine;  
KW docosahexenoic acid production; eicosapentenoic acid production.  
XX  
OS Shewanella putrefaciens.  
XX  
PN WO200042195-A2.  
XX  
PD 20-JUL-2000.  
XX  
PF 14-JAN-2000; 2000WO-US00956.  
XX  
PR 14-JAN-1999; 99US-0231899.  
XX  
PA (CALJ ) CALGENE LLC.  
XX  
PI Facciotti D, Metz JG, Lassner M;  
XX  
DR WPI; 2000-476063/41.  
XX  
PT New DNA sequences encoding for polyketide (PK)-like synthesis pathway  
PT genes from Shewanella, Vibrio and Schizochtrium, useful for creating  
PT transgenic plants that express poly-unsaturated long chain fatty acids  
XX  
PS Example 1; Fig 4E; 302pp; English.  
XX  
CC This invention describes novel DNA sequences encoding for polyketide  
CC (PK)-like synthesis (PKS-like) pathway genes from Shewanella, Vibrio  
CC and Schizochtrium. The nucleic acids are useful for isolating related  
CC molecules or in methods to detect organisms expressing the PKS-like  
CC genes. They are also useful for creating transgenic plants that express  
CC poly-unsaturated long chain fatty acids. The poly-unsaturated long chain  
CC fatty acids produced recombinantly are useful as dietary supplements for  
CC patients undergoing intravenous feeding or for preventing or treating  
CC malnutrition. The poly-unsaturated long chain fatty acids can also be  
CC incorporated into cooking oils, fats or margarine formulated so that in  
CC normal use the recipient receives a desired amount of poly-unsaturated  
CC long chain fatty acids. The nucleic acids are also useful in large scale  
CC production of docosahexenoic acid and eicosapentenoic acid, and for the  
CC modification of the fatty acid profile of host cells and edible plant  
CC tissues and/or plant parts. Transgenic production of polyunsaturated  
CC fatty acids in particular host cells allows quicker purification from  
CC natural sources such as fish or plants. This sequence represents the  
CC Shewanella putrefaciens PKS protein cluster ORF4 which is described in  
CC the method of the invention.

XX Sequence 970 AA;  
Query Match 7.2%; Score 147.5; DB 21; Length 970;  
Best Local Similarity 23.3%; Pred. No. 0.0024;  
Matches 53; Conservative 51; Mismatches 72; Indels 51; Gaps 10;

QY 15 LSVAVTQOLYAOQPNESLPVLEPVVITIDKSGMALANRITQMPHTTKVIVEEQIOEQAT 74  
DB 21 LTASLAMPVFAB--ETAAEQIERVAVT-----GSRIAKAELTOPAPVVSLSAEELTK--F 72  
QY 75 GSRQLADVMAQLIPSLGV-----SSGTTSNFGQTMHGRQVQFPLNG-----VPL 118  
DB 73 GNQDLGSLAEL-PAIGATNTIIGNNNSAGVSSADLRLRIGANRTLVLVNGKRYVAGQ 131  
QY 119 TGSRIISRLNSINPNQAVRIEVL-SGATSYGSGATGGLINIVTKSDLEEQFETRIGV 177

Db 132 PCSAEV--DLSTIPTSMISRVEIVTGGASAIYGSDAVSGVINVLKEDFEGFENAR--- 186  
QY 178 HGSKLSSEGIGYQVQSQSVAGVSENGNVLARLDVDYRTTGGAFDANGK 224  
DB 187 --TSGSTESVGTQ-----EHSFDILGGANVADGR 213  
RESULT 16  
AAB07697  
ID AAB07697 standard; Protein; 722 AA.  
XX  
AC AAB07697;  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE A Neisseria meningitidis BASB053 polypeptide.  
XX  
KW BASB053; Neisseria meningitidis infection; vaccine.  
XX  
OS Neisseria meningitidis.  
XX  
PN WO200042193-A1.  
XX  
PD 20-JUL-2000.  
XX  
PF 10-JAN-2000; 2000WO-EP00137.  
XX  
PR 15-JAN-1999; 99GB-0000959.  
XX  
PR 28-JAN-1999; 99GB-0001903.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX  
PI Ruelle J;  
XX  
DR WPI; 2000-476062/41.  
XX  
DR N-PSDB; AAAS9216.  
XX  
PT New Neisseria meningitidis polypeptide useful for diagnosis of  
PT Neisseria infection and for development of vaccines against such  
PT infection -  
PS Claim 3; Page 55-56; 92pp; English.  
XX  
CC The present sequence represents a Neisseria meningitidis BASB053  
CC polypeptide. The BASB053 polypeptide, or an antibody immunospecific  
CC for BASB053 may be identified in a biological sample in order to  
CC diagnose a Neisseria meningitidis infection in an animal. The BASB053  
CC polypeptides and polynucleotides may be used as vaccines, for  
CC generating an immune response in an animal. A composition comprising  
CC at least one antibody immunospecific for BASB053 may be used to  
CC treat humans infected with Neisseria meningitidis.  
XX  
SQ Sequence 722 AA;  
Query Match 7.2%; Score 146.5; DB 21; Length 722;  
Best Local Similarity 22.2%; Pred. No. 0.0019;  
Matches 88; Conservative 67; Mismatches 150; Indels 91; Gaps 23;

QY 18 AVTOOLYAOQPNESLPVLEPV-----VITIDKSGMALANRIT-----QMPHTTKV 63  
DB 20 ALSSSVFAAQATADLETVHIKQORSYNAIVTEKNGDYSGFAVTGKTIPASLREIPQSVSI 79  
QY 64 IYEQIQOQATGSRQLADVMAQLIPSLGVSGTTSNFGQTMHGRQV-QFLNGVPLTGSR 122  
DB 80 ITNQVKDRNVDY---FDQLARKTFGLRLVSLNDGGRSVYARGYSEYNDGLP----- 131  
QY 123 DISRLNSIN---PNQVA--RIEVLSGATSIY-GSGATGGLINIVTKSDLE----- 167  
DB 132 ---ACMQSINGTLNLFAPDRVEVMRGPGLFDSSGSGMGIVNLVRKPTKAFQGHAAAG 188  
QY 168 ---EQQFETRIGVHSGSKLSSEG-----IGYQVGOSVAGVSENGN-----VLARLDVDYR 213

Db 189 FGTGKQYKAEADVSGS-LNSDSVGRVWAQTVGASPPRAEKKNNRHETFAADMDINDP 247  
 QY 214 TT-GGAFDANGKRIAPD---PAQ-TDKODSKSLSVNTNVDNQ--LDDKONINLALTH---263  
 Db 248 TVLGAGYLVQGRRLAPYNGLPADANNKLPSPQHVFVGADMMKFKKNSHDFADLKHVFG 307  
 QY 264 -----YNDKO-DTDYAPDYGKRLAVLF-GEKPSLNAIK-----GLSLSEOPKTTKS 307  
 Db 308 NGCYGKVGKRYSDRDAADSNYA-FAGSKLGKTPAGRGCTADDKACAVGLGTEIKOKAL 366  
 QY 308 TENINHHDDLWGNNTIN---TNAYYR-----EKGR 335  
 Db 367 AFDASYSRPFRLGNTANEFVIGADYNNFRSTNEQGR 402  
 RESULT 17  
 AA75566  
 ID AA75566 standard; Protein; 708 AA.  
 AC AA75566;  
 XX  
 DT 21-MAR-2000 (first entry)  
 XX  
 DE Neisseria meningitidis ORF 760 protein sequence SEQ ID NO:2606.  
 XX  
 KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
 XX antibacterial; gene therapy.  
 OS Neisseria meningitidis.  
 XX  
 PN WO957280-A2.  
 XX  
 PD 11-NOV-1999.  
 XX  
 PF 30-APR-1999; 99WO-US09346.  
 XX  
 PR 01-MAY-1998; 98US-0083758.  
 PR 31-JUL-1998; 98US-0094869.  
 PR 02-SEP-1998; 98US-0098994.  
 PR 02-SEP-1998; 98US-0099062.  
 PR 09-OCT-1998; 98US-0103749.  
 PR 09-OCT-1998; 98US-0103794.  
 PR 09-OCT-1998; 98US-0103796.  
 PR 25-FEB-1999; 99US-0121528.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 XX  
 PI Frazer C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;  
 PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
 PI Tettelin H, Venter UC;  
 DR MPI; 2000-062150/05.  
 DR N-PSDB; AA254328.  
 XX  
 PT Novel Neisserial polypeptides predicted to be useful antigens for  
 PT vaccines and diagnostics -  
 XX  
 PS Claim 2; Page 1235; 1453pp; English.  
 XX  
 CC AA253015 to AA254536, AA254577 to AA254615, and AA774253 to AA775941  
 CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides  
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA255473 represent  
 CC PCR primers used in the exemplification of the present invention. The  
 CC polypeptides, the polynucleotides, antibodies and compositions of  
 CC the invention can be used as vaccines, as diagnostic reagents, and as  
 CC immunogenic compositions. The polypeptides can be used in the  
 CC manufacture of medicaments for treating or preventing infection due to  
 CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the  
 CC presence of Neisseria bacteria, or to raise antibodies. They may also  
 CC be used to screen for agonists or antagonists, which may themselves  
 CC have use as antibacterial agents. The polynucleotides of the invention

CC may also be used in gene therapy protocols.  
 XX  
 SQ Sequence 708 AA;  
 Query Match 7.1%; Score 146; DB 21; Length 708;  
 Best Local Similarity 21.8%; Pred. No. 0.0021;  
 Matches 96; Conservative 64; Mismatches 156; Indels 124; Gaps 25;  
 QY 18 AVTQOLVYAPNSLPTVELEPV-----VITIDKS-----GMALANRITQMPHTTKV 63  
 Db 20 ALSSVFPAQGTGLETVHKKQSRVNAITKQNGDYSSPAATVGTKIPASLREIPQSVSI 79  
 QY 64 IYEEQIQEATGSRQLADYMAQLIPSLGVSSGTTSNFGQTMHGRQV-OPLNGVPLTGSR 122  
 Db 80 ITNQGVKDRVDT---FDQLARTPELRVLSNDDGRSSVYARGVEYSEVINDGLP-----131  
 QY 123 DISRQLNSIN---PNOVA-RIEVLSGATSIY-GSATGGLINITYKSLF-----167  
 Db 132 ---AQMOSINGTLPLNFADRVKVRGSPGLPDSGEMGIVNLVAKRPTKAFQGHAAAG 188  
 QY 168 ---EEQFETRIGVHSGKLSSEG-----IGYQVQSVAGVSENGN-----VLARLDVDR 213  
 Db 189 FGTGKQYKAEADVSGS-LNSDSVGRVWAQTVGASPPRAEKKNNRHETFAADMDINDP 247  
 QY 214 TT-GGAFDANGKRIAPD---PAQ-TDKODSKSLSVNTNVDNQ--LDDKONINLALTH---264  
 Db 248 TVLGAGYLVQGRRLAPYNGLPADANNKLPSPQHVFVGADMMKFKKNSHDFADLKHVFG 307  
 QY 265 -----NDKODTYAPDYGKRLAVLFGEKPSLN-----AIKGLSLSEOPKTTKS 307  
 Db 308 NGCYGKVGKRYSDRDAADSNY-----TFAGSK--LNNTOADVAAGLGTDIKOKA--F 354  
 QY 308 TENINHHDDLWGNNTIN---TNAYYR-----EKGRFPVFAVPFSIAKALPILQSNWLP 359  
 Db 355 AFDASYSRPFRLGNTANEFVIGADYNNFRSTNEQGR-----STLSKS 396  
 QY 360 ATLDAVTKAPQARAVGVLS 379  
 Db 397 VALDGRFALPY--NGILQN 413  
 RESULT 18  
 ABG17885  
 ID ABG17885 standard; Protein; 226 AA.  
 AC ABG17885;  
 XX  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #17876.  
 XX  
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSB-) HYSB INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR MPI; 2001-639362/73.  
 DR N-PSDB; AAS82072.  
 PT New isolated polynucleotide and encoded polypeptides, useful in





```

DE A Neisseria meningitidis BASB053 polypeptide.
XX
XX BASB053; Neisseria meningitidis infection; vaccine.
XX
XX Neisseria meningitidis.
XX
XX WO200042193-A1.
XX
XX 20-JUL-2000.
XX
XX 10-JAN-2000; 2000WO-EP00137.
XX
XX 15-JAN-1999; 99GB-0000959.
XX
XX 28-JAN-1999; 99GB-0001903.
XX
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Ruelle J;
XX
XX WPI; 2000-476062/41.
XX
XX N-PSDB; AAA59217.
XX
XX New Neisseria meningitidis polypeptide useful for diagnosis of
XX Neisseria infection and for development of vaccines against such
XX infection.
XX
XX Claim 3; Page 56; 92pp; English.
XX
XX The present sequence represents a Neisseria meningitidis BASB053
XX polypeptide. The BASB053 polypeptide, or an antibody immunospecific
XX for BASB053 may be identified in a biological sample in order to
XX diagnose a Neisseria meningitidis infection in an animal. The BASB053
XX polypeptides and polynucleotides may be used as vaccines, for
XX generating an immune response in an animal. A composition comprising
XX at least one antibody immunospecific for BASB053 may be used to
XX treat humans infected with Neisseria meningitidis.
XX
XX Sequence 691 AA;
SQ
Query Match 6.8%; Score 139.5; DB 21; Length 691;
Best Local Similarity 22.7%; Pred. No. 0.0069;
Matches 80; Conservative 59; Mismatches 137; Indels 77; Gaps 21;
QY 47 GMAIANRITQMPHTTKVYEEQIOEQATGSRQADVAQILPSIGVSSGTTNFGQTMG 106
DB 32 GTXIPASLRIRIPQSVSITITQVKNRVDV---FDQLARTPTGLRLVLSNDRSSVYARG 88
QY 107 RQV-QEFLNGVPLTGSRRDISROLNSIN---PNOVA-RIVLSGANSIY-GSGATGGLIN 159
DB 89 YEYSEVNIIDGLP-----AQWQSIINGTLPNLFAFDRVEMRGPSGLPDSGEMGGINV 140
QY 160 IYVKSULE-----EBOFETRIGVHSGKLSSEG-----IGYVGQSVAGVSEN 201
DB 141 LVKRPRTKAFQGAAGFGTHKQYKADVSGS-LNSDSVGRGRVMAQTVGASPPRAEK 199
QY 202 GN-----VLARLDVRYTT-GGAFDANGKRIAE---PAQ-TDKODSKLSVNTNVDMQ- 250
DB 200 NRHETFPAAADWQINPPTVAGAGLVYQGRHLAPNGRLPAAANNKLBPLPHVIVGADPNK 259
QY 251 -LDDKONINIALTH-----YNDKQ-DTYADPYDGNRLAVLF-GEKPSLNAIK 294
DB 260 FKNNSHDVFPADLKHFGNGGYGVKGMKYSRDRDASNVA-PAGSKLGKMTDAGRPQCTAD 318
QY 295 ----GLISLQRPRTKSTFNINHYHDDLWGNITN---TNAYYR-----EKGR 335
DB 319 DKACAVGLGTEIKQKALAFDASYSRPRILCNTANEFVIGADYNRFRSTNBOGR 371

```

```

RESULT 21
AAAY43381
XX AAAY43381 standard; Protein; 1074 AA.
XX
XX AC AAAY43381;

```

```

XX
XX 26-JAN-2000 (first entry)
XX
XX M. catarrhalis strain 4223 tbpA protein.
XX
XX TbpB gene; Tbp2; transferrin binding protein; diagnosis; otitis media;
XX genetic immunisation; Moraxella infection; antigen; vaccine; detection;
XX antitumour antibody production; therapy.
XX
XX Moraxella catarrhalis.
XX
XX WO9952947-A2.
XX
XX 21-OCT-1999.
XX
XX 12-APR-1999; 99WO-CA00307.
XX
XX 14-APR-1998; 98US-0059584.
XX
XX (CONN-) CONNAUGHT LAB LTD.
XX
XX Myers LB, Schryvers AB, Harkness RE, Loosemore SM, Du R, Yang Y;
XX Klein MH;
XX WPI; 1999-620376/53.
XX
XX N-PSDB; AA231949.
XX
XX Nucleic acid encoding transferrin binding protein 2 of Moraxella
XX catarrhalis, useful for diagnostics, immunization and recombinant
XX protein production.
XX
XX Example 6; Fig 8; 114pp; English.
XX
XX This sequence is the Moraxella catarrhalis strain 4223 transferrin
XX binding protein (TbpA) of the invention. The DNA sequence is also
XX referred to as the TbpB gene. The TbpB gene is used to produce
XX recombinant Tbp2; for identification or diagnosis of Moraxella, or for
XX cloning related species, using hybridisation assays; and for genetic
XX immunisation against Moraxella infections, e.g. otitis media. The Tbp2
XX proteins are useful as antigens, either in vaccines (including components
XX of conjugate vaccines that contain antigens from other bacteria or from
XX tumours, in which case they elicit production of antitumour antibodies
XX that may be coupled to chemotherapeutic agents or biologically active
XX agents) or to raise antibodies (for use as diagnostic reagents and for
XX treating Moraxella infections), also for detecting Moraxella antibodies.
XX
XX Sequence 1074 AA;
SQ
Query Match 6.7%; Score 138; DB 20; Length 1074;
Best Local Similarity 23.6%; Pred. No. 0.017;
Matches 95; Conservative 48; Mismatches 149; Indels 110; Gaps 21;
QY 10 LSLPLISVA-----VTQQLVAQPNESLPTBLEPVITIDSGMALNRIQMPHTTKV 63
DB 22 LSLGLINITQVALNTADKAETDXTNLVVLVLDDEVVYAKKARK-ANEVGLGKRVYKT 80
QY 64 IYEEQIOEQATGSRQADVAQILPSIGS-VSSGTTNFGQTMG---RQVQFLNNGV---- 116
DB 81 A-BTINKEQVILNRDL---TRYDPEIAYVEQGRSSSGYSIRGDKNVAVLVDDINQA 135
QY 117 -----PLTGSR-DISROLNSINPQVARIIVLSGA-TSIVSGATGGLINIVTGS--- 164
DB 136 QHYALQVPAGKQVYAAGAINIEIYENNVASVELSKANSBVSQGLSSVAFVTTTAD 195
QY 165 ---DLEBEOFETR-----IGVHSGKLSSEGI-----GYVG 192
DB 196 IIRGDMQGVQTKTAYASIKNAMVNSVAAAGRAQSPSGILITDRGQRYKAHDDAYQGS 255
QY 193 QS-----VAGVSENGNV-----LARDVDVRYTGAFAFANG-RR 225
DB 256 QSFDRAVVATTPNNRTFTLIANECANGNYEACAAAGGQTKQAKPTNVRDKNVYKDYGPNR 315
QY 226 IAPPAQTDKODSKLSVNTNVDMQDLDRKONIN--LALTHYNDKQDTPDYADPYGNRLAVL 283

```

Db 316 LIPNPL---TQSKSLLLRPG---YQNDKHVGGVTEITKQNYAQDKTVPAY---LAVH 367  
Qy 284 FGKPSLN-----AIKGLSLSEQPKTT---KSTFNINHH 315  
Db 368 DIEKRLSNHAQGYQGNNGIRIDTIGPDSGVGINYAH 409

RESULT 22  
ID AAR75366 standard; Protein; 696 AA.  
XX AAR75366;  
AC AAR75366;  
XX 13-MAY-1996 (first entry)  
XX Adhesin.  
XX adhesin; plasmid pear; vector; vaccine; intestine colonisation.  
XX Escherichia coli.  
XX Key Location/Qualifiers  
FT Misc-difference 293 /note= "deduced residue from nucleotide sequence  
FT is His, this differs from Seq IDS, residue  
FT is Gly"  
XX WO9600233-A1.  
XX 04-JAN-1996.  
XX 07-JUN-1995; 95WO-US06994.  
XX 24-JUN-1994; 94US-0265714.  
XX (CHIL-) CHILDREN'S HOSPITAL & MEDICAL CENT.  
XX (UNIW ) UNIV WASHINGTON.  
XX (UNIW ) UNIV WASHINGTON STATE RES FOUND.  
XX Besser TE, Bilge SS, Tarr PI, Vary JC;  
XX WPI; 1996-068826/07.  
XX N-PSDB; AAT10105.  
XX Chromosomal DNA from E. coli O157:H7 encoding epithelial adhesin -  
XX isolated on plasmid pSC (overlap), for use as a vaccine to prevent  
XX bacterial colonisation of bovine intestine  
XX Disclosure; Page 31-32; 42pp; English.  
XX A candidate adhesin (AAR75366) was identified that is a homologue  
XX of the IrgA protein of Vibrio cholerae. The adhesin enables  
XX Escherichia coli O157:H7, an antibiotic-resistant, virulent and  
XX common food-borne pathogen, to adhere to epithelial cells.  
XX Recombinant adhesin was obcd. by expression of the encoding sequence  
XX (see AAT10105) in E. coli HB101 (pear). The adhesin can be used as a  
XX vaccine for immunisation of cattle against disease or colonisation  
XX of mucosal surfaces by O157:H7, thus increasing the safety of food  
XX derived from cattle.  
XX Sequence 696 AA;  
XX Query Match 6.6%; Score 135.5; DB 17; Length 696;  
XX Best Local Similarity 22.4%; Pred. No. 0.015;  
XX Matches 57; Conservative 43; Mismatches 82; Indels 73; Gaps 12;  
Qy 40 VTIIDKSGMALANRITOMPHHTKYIYEBIQICQATGSRQLADVMAQL--IPSLGVSSGTT 97  
Db 27 VMIVASAG--YEKLTNAAASVSVISQEBLQ-----SSQYHDLAALRSVGVGVESGTG 79  
Qy 98 SNFG-----QTMHGROVQPLNGVPLTGTGRDIS-RQLNSINPN-----QVARIIVLSGA 145

Db 80 KTGGLTISIRGMPASPYTLILIDGVROGGSSDVTNPGFSAMNTGMPPLAAIERIEVIRGP 139  
Qy 146 TS-YGSGATGGLINIVTKSLDEEQPETRIGVH-----GSKLSSEGI 187  
Db 140 MSTLYGSDAMGGVVNIITRKNADKWLSSVNAQLNESKNKGSSQPNFWSSGPLVDDSV 199  
Qy 188 GYQV-----GOSVAGVSENGVNLARLDVDTYTTGAFDANGKRIAPBPAQTDKQDSK 239  
Db 200 SLQVRGSTQQRQSGSSVTSLS-----DTAGTRI--PYPT-----ESQ 233  
Qy 240 SLSVNTNVDMQDDK 254  
Db 234 NYNLGARLDWKASEQ 248

RESULT 23  
ID ABG24651 standard; Protein; 905 AA.  
XX ABG24651;  
AC ABG24651;  
XX 18-FEB-2002 (first entry)  
XX Novel human diagnostic protein #24642.  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
XX Homo sapiens.  
XX WO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US08631.  
XX 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
XX N-PSDB; AAS88838.  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity -  
XX Claim 20; SEQ ID No 55010; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. ABG00010-ABG30377 represent novel human  
XX diagnostic amino acid sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 905 AA;

Query Match 6.6%; Score 135; DB 22; Length 905;

Best Local Similarity 21.9%; Pred. No. 0.024; Mismatches 105; Indels 66; Gaps 14;

Matches 63; Conservative 54; Mismatches 105; Indels 66; Gaps 14;

14 LLSAVATQQLVAQPNESLPVLEPVITIDKSGMALNRITQMPHTTKVIEBQIOEA 73

283 LVNLGIYGAQAQEPPTDPVSHDDTIVT-----AAEQNLQAPGVSTTAD-IRKNP 334

74 TGSROLADVAQQLIPSLGVSGTSGNFGQTMGRQVQ-----FLNGVPLT----- 119

335 V-ARDVSKII-RTMPGVNLGNSTG--GGRGNRRQIDIRMGPEPNTLLIDKRPVSSRNS 390

120 -----GSRDISRQLNSINPNQVARIIEVLG-ATSIYSGATGGLINTVTKSDLEEQFB 172

391 VRQGWGRGERTDRTGTSWPEPMIERIEVLKRGPSARYGNAAGVNNITTKGSGE---- 446

173 TRGVHGS-----KLSSEGIQVGVQSVAGV---SENGVNLARLDVYRTTGAFDA 221

447 ----WHSWDAYFNAPENHEGATKRTNPSLTGPDGDFSPFLYGNLDKQVTRGISTRA 502

222 NGKRI--APEPAQTDKQDSKLSVNTNV--DMQLDDKONIMALTTHY 264

503 ISPRVPERMPRRYQPARRGQPFHHTKILGKWTI-----FAHY 541

RESULT 24

ABG28872 standard; Protein; 908 AA.

AC ABG28872;

XX 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #28863.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

XX WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS93059.

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity -

XX Claim 20; SEQ ID No 59231; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 908 AA;

Query Match 6.6%; Score 135; DB 22; Length 908;

Best Local Similarity 21.9%; Pred. No. 0.024;

Matches 63; Conservative 54; Mismatches 105; Indels 66; Gaps 14;

14 LLSAVATQQLVAQPNESLPVLEPVITIDKSGMALNRITQMPHTTKVIEBQIOEA 73

286 LVNLGIYGAQAQEPPTDPVSHDDTIVT-----AAEQNLQAPGVSTTAD-IRKNP 337

74 TGSROLADVAQQLIPSLGVSGTSGNFGQTMGRQVQ-----FLNGVPLT----- 119

338 V-ARDVSKII-RTMPGVNLGNSTG--GGRGNRRQIDIRMGPEPNTLLIDKRPVSSRNS 393

120 -----GSRDISRQLNSINPNQVARIIEVLG-ATSIYSGATGGLINTVTKSDLEEQFB 172

394 VRQGWGRGERTDRTGTSWPEPMIERIEVLKRGPSARYGNAAGVNNITTKGSGE---- 449

173 TRGVHGS-----KLSSEGIQVGVQSVAGV---SENGVNLARLDVYRTTGAFDA 221

450 ----WHSWDAYFNAPENHEGATKRTNPSLTGPDGDFSPFLYGNLDKQVTRGISTRA 505

222 NGKRI--APEPAQTDKQDSKLSVNTNV--DMQLDDKONIMALTTHY 264

506 ISPRVPERMPRRYQPARRGQPFHHTKILGKWTI-----FAHY 544

RESULT 25

AAU33610 standard; Protein; 815 AA.

AC AAU33610;

XX 14-FEB-2002 (first entry)

DE Pseudomonas aeruginosa cellular proliferation protein #54.

KW Antisense; prokaryotic cellular proliferation protein;

XX antibiotic; antibacterial; drug design.

OS Pseudomonas aeruginosa.

XX WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

XX 23-MAY-2000; 2000US-206848P.

XX 26-MAY-2000; 2000US-207272P.

XX 23-OCT-2000; 2000US-242578P.

XX 27-NOV-2000; 2000US-253625P.

XX 22-DEC-2000; 2000US-257931P.

XX 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen XL, Zyskind JW, Wall D, Trawick JD, Carr GJ;  
 PI Yamamoto RT, Xu HH;  
 XX WPI; 2001-611495/70.  
 DR N-PSDB; AAS51469.  
 XX New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX Example 3; Seq ID No 5106; 511pp; English.  
 XX The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX SQ Sequence 815 AA;

Query Match 6.5%; Score 134; DB 22; Length 815;  
 Best Local Similarity 21.7%; Pred. No. 0.025;  
 Matches 80; Conservative 63; Mismatches 166; Indels 60; Gaps 16;  
 QY 33 TVELEPVVITDKSG-----MALANRIT---QMPHTTKVIYERIQEQATG 75  
 DB 124 SVDLGATMITSNQLGNTEDSGSYTPGTATATATRLVLTPTREPQSTVTVTRQNMDD--FG 181  
 QY 76 SQQLADVMAQLIPSLGVSGTTSNFGQTHGQV-QFLNGVPLTGSRIQL-NSINP 133  
 DB 182 LNNIDDMVRH-TPGITVSAYDTRNNYARGFSINNFOYDGPST-ARNVGYSGNLTSD 239  
 QY 134 NQV-ARIEVLGATSIY-GSGATGGLINVTKSDLSEEQFETRIGVHGSKLSEGIQV 191  
 DB 240 MAIYDRVEVLKGTGLTGGSLGATINILIRKPTHEFGHVELG-----AGSDNYRS 293  
 QY 192 GOSVAG-VSENGNVLARLDVYRTTGGAFDANGKRIAPEPAQTDKQSKSLSVNTNVDWQ 250  
 DB 294 ELDSVGLTESGNVRGRAVAQDKHSFMD-----HYERKTSVYGI--- 335  
 QY 251 LDDQNINLALHYNDKQTDVADYGNRLAVLFGKPSLNAI-----KGLSLSEOPKTT 305  
 DB 336 LEFDLPNPTMLTVGADYQDNPKSGWGSFPFLDSQGNRNDVSRFNNGAKWSSWEQVT 395  
 QY 306 KSTF-NINHHDDLMGNTI-----NTNAYYREKGRYPVPVAPFSAIKALPILQSMNLPSA 360  
 DB 396 RTVFANLEHNFANGWGVKQVOLDHKINGYHAPLGAIMGDPAPDNSAKIYVAKYGTETKN 455  
 QY 361 TLDAYTKAP 369  
 DB 456 SLDIYLTGP 464

RESULT 26  
 ID AAW35312  
 XX AAW35312 standard; Protein; 1074 AA.  
 AC AAW35312;  
 XX

DT 14-APR-1998 (first entry)  
 XX M. catarrhalis 4223 transferrin binding protein tbpA.  
 DE Transferrin binding protein; tbpA; immunogen; vaccine; protection;  
 XX otitis media; antibody; diagnosis; therapy; carrier;  
 KW Gene isolation.  
 KW Moraxella catarrhalis.  
 OS WO9732980-A1.  
 XX 12-SEP-1997.  
 XX 07-MAR-1997; 97WO-CA00163.  
 XX 03-JAN-1997; 97US-0778570.  
 PR 08-MAR-1996; 96US-0613009.  
 XX (CONN-) CONNAUGHT LAB LTD.  
 XX Du R, Harkness RE, Klein MH, Loosmore SM, Myers LE;  
 PI Schryvers AB, Yang Y;  
 XX WPI; 1997-457533/42.  
 DR N-PSDB; AAT95247.  
 XX DNA encoding transferrin receptor of a Moraxella strain - also  
 PT proteins, useful in vaccines, as diagnostic agents and in the  
 PT production of antibodies  
 XX Claim 6; Fig 5; 162pp; English.  
 PS The present sequence is the Moraxella catarrhalis 4223  
 XX transferrin binding protein tbpA, which can be used as an  
 CC immunogen, e.g. in vaccines to protect against diseases caused  
 CC by M. catarrhalis (specifically otitis media), or to raise  
 CC antibodies for diagnosis and therapy. It can also be used as a  
 CC carrier for other antigenic determinants, e.g. of bacteria  
 CC containing polysaccharide antigens or abnormal polysaccharides  
 CC present on tumour cells, particularly to make conjugate vaccines.  
 CC The tbpA DNA can be used to detect nucleic acid encoding  
 CC transferrin receptor protein, e.g. for diagnosis or gene  
 CC isolation, by usual hybridisation assays.  
 XX Sequence 1074 AA;

Query Match 6.5%; Score 134; DB 18; Length 1074;  
 Best Local Similarity 23.4%; Pred. No. 0.037;  
 Matches 94; Conservative 48; Mismatches 150; Indels 110; Gaps 21;  
 QY 10 LSLPILLSVA-----VTQQLYAPQNESLPTVELEPVVITDKSGMALANRITQMPHTTKV 63  
 DB 22 LSLGLLITQVALANTTADKAEATDKNLVVLDVTVTAKNKARK-ANEVTGLGKVVKT 80  
 QY 64 IYEQIQOATGSRQOLADVMAQLIPSLG-VSSGTTSNFGQTHG---RQVQFLNGV--- 116  
 DB 81 A-ETINKEQVLNIRD-----TRYDPIAVVEQGRGASSGYSIRGMDKNRVAVLVDGINQA 135  
 QY 117 -----PLTGSR-DISQLNSINPQVARIELVLSGA-TSIYGSGATGLINIVTKS--- 164  
 DB 136 QHYALQGVAGKNYAAGGAINIEIYENVRSEISKGANSSEYSGSALSGSAFVTKTADD 195  
 QY 165 ---DLEEEQFETR-----IGVHGSKLSSSEGI-----GYQVG 192  
 DB 196 IIKGDKWGVQTKTAYASKNNWVNSVAAAGKAGSFGSLIITDRRQGEYKAHDDAYQGS 255  
 QY 193 QS-----VAGVSENGV-----LARLDVYRTTGGAFDANG-KR 225  
 DB 256 QSFDRAVATDPNNRTFLIANECANGNYEACAAGQOTKLOAKPTNVRDKNVNVTGPNR 315  
 QY 226 IAPEPAQTDKQSKSLSVNTNVDWQDKQININ--LALTHYNDKODTDYAPDYGRLAVL 283

Db 316 LIPNPL---TQDSKSLLRPG--YQLNDKHVYGVVETITQNTAMQDKTVPAY---LTVH 367  
 Qy 284 FGEKPSLN-----AIKGLSLSEQPKT---KSTFNINATH 315  
 Db 368 DIKSRSLSNHQAQNGYQGNLIGRIDTIGPDSGYINATH 409

RESULT 27  
 AAU87504  
 ID \*AAU87504 standard; Protein; 228 AA.  
 XX  
 AC AAU87504;  
 XX  
 DT 05-JUN-2002 (first entry)  
 XX  
 DE Novel central nervous system protein #14.  
 XX  
 KM Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;  
 KM hyperproliferative disorder; neoplasm; cardiovascular disorder;  
 KM cardiac arrest; cerebrovascular disorder; ischaemia; angiodysplasia;  
 KM nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;  
 KM acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;  
 KM adenocarcinoma; reproductive system disorder; testicular feminisation;  
 KM endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;  
 KM respiratory disorder; renal disorder; kidney failure; blood disorder;  
 KM myocardial infarction; wound healing; cell proliferation; skin aging;  
 KM food additive; food preservative; gene therapy.  
 OS Homo sapiens.  
 XX  
 XX  
 PN WO20015318-A2.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01332.  
 XX  
 XX 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226868.  
 PR 23-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239835.  
 PR 13-OCT-2000; 2000US-0239837.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.









PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI; 2001-581633/65.  
 DR N-PSDB; ABRK3522.  
 XX  
 PT New isolated nucleic acid encoding a protein for diagnosing,  
 PT preventing, treating or ameliorating medical conditions and used as  
 PT food additives or preservatives -  
 XX  
 PS Claim 9; SEQ ID No 710; 837pp; English.  
 XX  
 CC The invention describes an isolated nucleic acid molecule (I) encoding a  
 CC novel central nervous system protein. (I) and polypeptides (II) encoded  
 CC by (I), are used to treat a medical conditions and in diagnosis of a  
 CC pathological condition. Disorders which are diagnosed or treated include  
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative  
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders  
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,  
 CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and  
 CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses  
 CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders  
 CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,  
 CC adenocarcinoma and irritable bowel syndrome, reproductive system  
 CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes  
 CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.  
 CC leukaemia, disorders involving neovascularisation e.g. malignancies,  
 CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.  
 CC acute kidney failure and blood related disorders e.g. myocardial  
 CC infarction. The polypeptides can also be used to aid wound healing and  
 CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
 CC maintain organs before transplantation, for supporting cell culture of  
 CC primary tissues, to regenerate tissues and in chemotaxis. The  
 CC polypeptides can also be used as a food additive or preservative to  
 CC increase or decrease storage capabilities, fat content, lipid, protein,

Query Match 6.5%; Score 133.5; DB 22; Length 252;  
 Best Local Similarity 25.0%; Pred. No. 0.0053;  
 Matches 47; Conservative 38; Mismatches 62; Indels 41; Gaps 9;  
 QY 14 LLSAVTQQLYVQPNESLPTELEPVVITIDSGMALAKRITQMHTTVITVEQIQEA 73  
 Db 13 LVNLGIYVGAQAEPTDVTFSHDDIVVT-----AEONLQAPGVSTITADE-IRKXP 64  
 QY 74 TGSROLADYMAOLIPELVSGSTSNFGOTMGRQVQ-----FLNCGVPLT----- 119  
 Db 65 V-ARDVSKII-RFMGVNLTGNST--GQRNNRQIDIRGMPEPTLLIDGRVSSRS 120  
 QY 120 -----GSPDISRQNLNSINPNQVARIYVLSG-ATSIYSGATGCLINITYKSDLREPOE 172  
 Db 121 VQGKMGKRGERTDGTDSWVPEMIBRIEVARGPAAARGMAGGVNITTKGSGS----- 176  
 QY 173 TRIGVHGS 180  
 Db 177 ---WHGS 180  
 RESULT 30  
 AAU17142  
 ID AAU17142 standard; Protein; 274 AA.  
 XX  
 AC AAU17142;  
 XX  
 DT 07-NOV-2001 (first entry)  
 XX  
 DE Novel signal transduction pathway protein, Seq ID 707.  
 XX  
 KW Neuroprotective; cytosolic; dermatological; immunosuppressive; tumour;  
 KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;  
 KW immune system disorder; rheumatoid arthritis; inflammatory condition;  
 KW organ transplant rejection; infection; hepatitis C; blood disorder;  
 KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;  
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;  
 KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;  
 KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;  
 KW reproductive system; gastrointestinal; liver disorder; AIDS;  
 KW acquired immune deficiency syndrome.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200154733-A1.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01312.  
 XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-020515.  
 PR 07-JUN-2000; 2000US-0204967.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.



CC T-cells, to induce higher affinity antibodies, and as a means to induce  
CC tumour proliferation in pathologies e.g. acquired immune deficiency  
CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction  
CC pathway protein, amino acid sequences of the invention.

Query Match 6.5%; Score 133.5; DB 22; Length 274;  
Best Local Similarity 25.0%; Pred. No. 0.006;  
Matches 47; Conservative 38; Mismatches 62; Indels 41; Gaps 9;

QY 14 LLSVATQQLYQNPESLPVLEPVIITIDKSGMALANRITQMPHTTKYIEQIQEQA 73  
DB 37 LVLGLGYGAQAQEPDTDPVSHDDTVVT-----AAEQNLQAPGVSTTTADE-IRKXP 88  
QY 74 TGSROLADVAQAOLIPSLGVSGTTSNPGQTMGRQVQ-----FLNGVPLT----- 119  
DB 89 V-ARDVSKII-RMPGVNLTGNST--GQGNKRQIDIRMGPENTLILDGKPVSSKRS 144  
QY 120 -----GSRDISRQLNSINNOVARIEVLG-ATSIYSGATGALINVTKSDLSEBOFE 172  
DB 145 VRQGMGERTRDGTSMVPEMERIEVLKGPAAARYGNAGGVNITIKKSGE----- 200  
QY 173 TRGVHGS 180  
DB 201 ---WEGS 204

## RESULT 31

AAW01461 standard; Protein; 715 AA.

AAW01461;

24-FEB-1997 (first entry)

NTHI HxuC protein.

HxuC; NTHI; vaccine; genetic immunisation; diagnosis;

meningitis; pneumonia; bacteraemia; otitis media.

OS Haemophilus influenzae nontypeable strain TN106.

Key Location/Qualifiers

6..18 /label= TonB box

19..21 /note= "signal peptidase I cleavage site"

27..34 /note= "motif region"

67..71 /note= "motif region"

93..103 /note= "motif region"

121..146 /note= "motif region"

505..514 /note= "motif region"

577..583 /note= "motif region"

691..703 /note= "motif region"

704..715 /note= "motif region"

WO9633275-A1.

24-OCT-1996.

15-APR-1996; 96WO-US05167.

20-APR-1995; 95US-0425843.

(TEXA ) UNIV TEXAS SYSTEM.

XX Cope LD, Hansen EJ, Hanson MS, Jarosik GP;  
PI MPI: 1996-485781/48.  
DR N-PSDB; AAT44519.

PT Genes encoding H. influenzae HxuC and HxUB surface-expressed  
PT protein(s) - useful in the prepn. of vaccines for children against  
PT H. influenzae infection

PS Claim 9; Page 131-136; 189pp; English.

CC The HxuC protein (AAW01461, see also AAW01462) of nontypeable  
CC Haemophilus influenzae (NTHI) strain TN106 is a 78 kDa outer  
CC membrane protein involved in the utilisation of low levels of free  
CC haem. It shows homology to TonB-dependent outer membrane proteins  
CC of other bacteria, and was identified from an open reading frame  
CC (see also AAT44519) located downstream of the TN106 HxUA gene.  
CC Another outer membrane protein, HxUB (AAW01464, see also AAW01465), was  
CC similarly identified. Recombinant HxuC and HxUB proteins can be  
CC prepnd. in transformed host cells and used: to prepare vaccines  
CC against NTHI infection; to raise diagnostic antibodies; and to  
CC prepare NTHI diagnostic or therapeutic compsns.

SQ Sequence 715 AA;

Query Match 6.5%; Score 133; DB 17; Length 715;

Best Local Similarity 23.8%; Pred. No. 0.025; Matches 84; Conservative 55; Mismatches 135; Indels 82; Gaps 18;

QY 7 FQMLSLPLSLVAVTQQLYQNPESLPVLEP--VITIDKSGMALANRITQMPHTTKVI 64  
DB 3 FQKSLAIIITTLVTNMAAQ-----SVELDSINVIATDPSRFAYT----- 43  
QY 65 YEBQIQEQATGSRQADVAQAOL-IPSLGVSGTTS-----NFGQTMGRQVQFLNGVP 117  
DB 44 -DEKOSKDSLKQATSVAALEDIPNVDRGSSRSIAQKPNIRGLSDRNVQ-VIDGV- 100  
QY 118 LQGRDISRQLNSINP-NOVARIEVLG-ATSIYSGATGALINVTKSDLSEBOFETRI 175  
DB 101 -RONPDLARGSYFLPMSLQIEIYVKGPSSSLWGSGLGVAVAKTPPALDLKKNDF 159  
QY 176 GVHGSKLSEGGIGYGVQSVAGVSENGVNLARDVDYRTTGAFPANGKRIAPPAQTDK 235  
DB 160 GV---KIRQ---GYQTANNLS--ERDASVPAAND-KFDVLISAFNNANL-----RTCK 205  
QY 236 QDS-----KSLSVNTNVQDLDDKQINILALTHYNDKQDTDYADPYGNRLAVLFGSEKPS 289  
DB 206 GKNLANTAVKQFGGLAKFGQINDARVELSHRETRFKQTA-----PS 248  
QY 290 LNAIKGLSSEQPKTKSTFNINHHDDLWG-----NTINNNAYYRREKGRF 336  
DB 249 NNEVENELTNE-----KIIDOINEFGSNNGLPORAKPSSETSAFYSKVKTRF 296

## RESULT 32

AAW98606 standard; Protein; 767 AA.

AAW98606;

31-MAR-1999 (first entry)

H. pylori GHPO 147 protein.

KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;  
KW peptic ulcer disease.

OS Helicobacter pylori.

WO9843478-A1.

08-OCT-1998.





```

Db 210 VFHNYDLQPSNDSMTSGTSEVNSVLAKINGYINETDSISVNLTRDSTNLLRPT 269
Qy 258 NLALTHYND--KQDTDYADYGNLAVLFGKPSLNAIKGLSLSEQPKTKTKSTFNINHH 315
Db 270 TSALSANDPSQAPFVIDGKELA-----HTINFHNHLSL-----KYKH 310
Qy 316 DDLWGNTIN-----TNAYYRREKGRFPFVAPPS 344
Db 311 EG--GPNENQPRVESTAFELVRGGNYNPVNPFA 342

RESULT 36
AAW35314
ID AAW35314 standard; Protein; 1070 AA.
XX
AC AAW35314;
XX
DE 14-APR-1998 (first entry)
XX
M. catarrhalis Q8 transferrin binding protein tbpA.
XX
Transferrin binding protein; tbpA; immunogen; vaccine; protection;
KW otitis media; antibody; diagnosis; therapy; carrier;
KW gene isolation.
XX
OS Moraxella catarrhalis.
XX
PN W09732980-A1.
XX
PD 12-SEP-1997.
XX
PF 07-MAR-1997; 97WO-CA00163.
XX
PR 03-JAN-1997; 97US-0778570.
PR 08-MAR-1996; 96US-0613009.
XX
PA (CONN-) CONNAUGHT LAB LTD.
PI Du R, Harkness RE, Klein MH, Loosmore SM, Myers LE;
PI Schryvers AB, Yang Y;
XX
DR WPI; 1997-457533/42.
DR N-PSDB; AAT95249.
XX
DNA encoding transferrin receptor of a Moraxella strain - also
PT proteins, useful in vaccines, as diagnostic agents and in the
PT production of antibodies
XX
PS Claim 6; Fig 10; 162pp; English.
XX
The present sequence is the Moraxella catarrhalis Q8
transferrin binding protein tbpA, which can be used as an
immunogen, e.g. in vaccines to protect against diseases caused
by M. catarrhalis (specifically otitis media), or to raise
antibodies for diagnosis and therapy. It can also be used as a
carrier for other antigenic determinants, e.g. of bacteria
containing polysaccharide antigens or abnormal polysaccharides
present on tumour cells, particularly to make conjugate vaccines.
CC The tbpA DNA can be used to detect nucleic acid encoding
transferrin receptor protein, e.g. for diagnosis or gene
isolation, by usual hybridisation assays.
XX
SQ Sequence 1070 AA;
Query Match 6.4%; Score 130; DB 18; Length 1070;
Best Local Similarity 23.1%; Pred. No. 0.078;
Matches 93; Conservative 48; Mismatches 151; Indels 110; Gaps 21;

```

```

Qy 10 LSLPILLSVA-----VTOOLYAQPNESLPTVELEPWITIDKSGMALANRITOMPHTTKV 63
Db 19 LSLGLLNTQVALANTTADKAEATDKTNLVVLDETIVTAKQARK-ANEVTGLGKVVKT 77

```

```

Qy 64 IYEQIOEQATGSRQLADVMAQLIPSLG-VSSGTTSNFGQTMHG---ROYQFLNGV--- 116
Db 78 A-ETINKEQVLNIRDL-----TRYDPGIAVVEQGRGASSGYSIRGMDKNRVAVLVDGINQA 132
Qy 117 -----PLTGSR-DISBQLSNPNNOVARIEVLVLSGA-TSIYSGATGGLINIVTKS--- 164
Db 133 QHYALQGFVAGKVAAGGAINETIYENVRVSEISKGANSSEYSGSALSGVAFVTKTADD 192
Qy 165 ---DLEEEQFETR-----IGVHGSKLSSEGI-----GYQVG 192
Db 193 IIKDKMGVQTKTAYASKKNAMVNSVAAAGKAGSPSGLIIYTDRRGQEKYKAHDDAYQGS 252
Qy 193 QS-----VAGVSENGV-----LARLDVVDYTTGGAFDANG-KR 225
Db 253 QSPDRAVATTPDNNPKFLIANECANGNYEACAAGGQTKLOAKPTNVRDKVNVKDYTGPNR 312
Qy 226 IAEPAQOTDKQDSKLSVNTNVDMDLQKNIN--LALTHYNDKQDQDYADYDYNRLAVL 283
Db 313 LIPNPL-----TQDSKSLLRPG--YQNDKHYVGVYEITKQNYAMQDKTVPAY---LTVH 364
Qy 284 FGEKPSLN-----AIKGLSLSEQPKT---TKSTFNINHH 315
Db 365 DIEKRLSNHGQANGYVQGNLGERIRDAICANSYGGINVAH 406

RESULT 37
AAW18061
ID AAW18061 standard; Protein; 930 AA.
XX
AC AAW18061;
XX
DE 08-SEP-1997 (first entry)
XX
Pasteurella haemolytica transferrin binding protein TbpA.
XX
Transferrin binding protein; TbpA; receptor; vaccine;
KW pasteurellosis; antibody.
XX
OS Pasteurella haemolytica serotype A1 strain Hi96.
XX
FH Key Location/Qualifiers
FT Peptide 1..28
FT /label= Sig_peptide
XX
PN W09720934-A1.
XX
PD 12-JUN-1997.
XX
PF 29-NOV-1996; 96WO-CA00791.
XX
PR 01-DEC-1995; 95US-0008569.
PR 01-DEC-1995; 95CA-2164274.
XX
(LORY/) LO R Y C.
PA (POTT/) POTTER A A.
PA (SCHR/) SCHRYVERS A B.
XX
PI Lo RYC, Potter AA, Schryvers AB;
XX
DR WPI; 1997-319780/29.
DR N-PSDB; AAT67235.
XX
Pasteurella haemolytica transferrin binding proteins - used in
PT vaccines for prophylaxis and treatment of infection caused by
PT Pasteurella spp.
XX
PS Claim 3; Page 84-86; 130pp; English.
XX
Transferrin binding proteins TbpA (AAW18061) and TbpB (AAW18062) of
CC Pasteurella haemolytica serotype A1 strain Hi96 are transferrin
CC receptors of approx. 100 kDa and 60 kDa, respectively. Their
CC amino acid sequences were deduced from isolated gene sequences
CC (AAT67235-36) and show regions of homology with the corresponding

```

CC *Neisseria meningitidis*, *N. gonorrhoeae*, *Haemophilus influenzae* and  
 CC *Actinobacillus pleuropneumoniae* Lbp and Tbp proteins. TbpA and  
 CC TbpB polypeptides can be expressed in transformed host cells for  
 CC use in vaccines that protect ruminants against diseases caused by  
 CC *P. haemolytica*, partic. bovine pneumonic pasteurellosis. They can  
 CC also be used to identify (ant)agonists and to raise antibodies  
 CC useful e.g. in passive immunisation.

XX Sequence 930 AA;

Query Match 6.3%; Score 129.5; DB 18; Length 930;  
 Best Local Similarity 26.0%; Pred. No. 0.071;  
 Matches 60; Conservative 30; Mismatches 108; Indels 33; Gaps 10;

QY 5 HYQWMLPLLSVAVTQOLYAPNESLPTVE-----LEEVITIDKSGMALNRITOMP 58  
 DB HHRFYSVALTVLPAHSYGAATENKKEENNDLAVLDEVITESHYAHQRONEVTLGLG 65  
 QY 59 HTTKVYIEQIQEQATGSRQADVMQALPSLG-VSSGTTSNFGQTMHG--ROYQPLIN 114  
 DB 66 KVKV-NYHEMSKQOILGIRDL---TRYDPGISVEEQRGASGVALRGVDKXKRVSLVD 120  
 QY 115 GVP-----LTGSRDISROLNSINPNQVARIENVLSGATSI-YGSGATGLINIVTK--SD 165  
 DB 121 GLPQASHYHTLGSDDANGAINEIENYINRSTELSKGASAEYSGAGALGPRTKDAQD 180  
 QY 166 LEEBOETRIQVH---GSKLSEGIQYQVQSVAGVSENGNVLARLDVDR 213  
 DB 181 IIRB-----GQHWGLDSKTSYASKSHPLQSLAAAGAGFALVATATR 225

RESULT 38  
 AAR95568  
 ID AAR95568 standard; Protein; 790 AA.

XX AAR95568;

DT 28-AUG-1996 (first entry)

XX *N. gonorrhoeae* B haemoglobin receptor.

KW Bacterial haemoglobin receptor; hmbR gene; antibody; vaccine;  
 KM gonorrhoea.

OS *Neisseria gonorrhoeae* strain MS11A.

PN MO9612020-A2.

PD 25-APR-1996.

PF 17-OCT-1995; 95WO-US13623.

PR 18-OCT-1994; 94US-0326670.

PA (UYOR-) UNIV OREGON HEALTH SCI.

PI Heffron F, Hwa V, Nassif X, So M, Stojiljkovic I;

DR WPI: 1996-222006/22.

DR N-PSDB; AAT27000.

PT DNA encoding *Neisseria* haemoglobin receptor proteins - for use in  
 PT preparing polypeptide(s) and antibodies for vaccines for e.g.  
 PT meningitis

PS Claim 5; Page 67-69; 104pp; English.

CC A bacterial haemoglobin receptor (AAR95568) was identified as the  
 CC product of the hmbR gene (AAT27000) of *N. gonorrhoeae* strain MS11A.  
 CC The haemoglobin receptor is thought to be involved in heme  
 CC utilisation by the bacterium. This dependence on host iron stores  
 CC is a potential route for therapeutic intervention strategies. The  
 CC receptor can be expressed in transformed hosts, e.g. attenuated

CC *Salmonella* cells. It is useful as a vaccine for gonorrhoea, or can  
 CC be used to raise antibodies of diagnostic appln. Similar receptor  
 CC proteins were obsd. from *N. meningitidis* serotypes A (AAR95566), B  
 CC (AAR95567) and C (AAR95565).

XX Sequence 790 AA;

Query Match 6.3%; Score 128; DB 17; Length 790;  
 Best Local Similarity 23.4%; Pred. No. 0.075;  
 Matches 95; Conservative 44; Mismatches 151; Indels 116; Gaps 21;

QY 12 LPLLSVAVTQOLYAP-----NESLPT-VELBHVITIDKSGMALNRITOMPHTTK 62  
 DB 7 LPL--AALVGSIFGNVPLADBAATETTPVKAIEKVRKVDQINAPATYERV-----NLG 59  
 QY 63 VYIEEQIQEQATGSRQADVMQALPSLGVSQTTSNFG---QTHGROYQFLANGVPLT 119  
 DB 60 RIQDEMIRDKLVRYSTDV-----GLSDSGRQKQFPAVAGVGNVGSIDGVSIF 111  
 QY 120 GSRDIS---ROLN-----SINPNQVARIENVLSGATSI-YGSGATGLINIVTKSD----- 165  
 DB 112 DSEENSLVARYGNFNFSRLSIDBELVRNIEIAKGADSFMTGSGALGGVNYOTLQGHDL 171  
 QY 166 LEEBOETRIQVHGSFLSSE---GIGYQV-----GQSVAGVSENGVNLAR 207  
 DB 172 LDRQFGVMMKNGYSRNHEMTWTLFGVSNDRVDALLYSQRRGHETESAGERG----- 226  
 QY 208 LDVDRYTTG---GAFNANKRIAPAPQOTDKODSKLSVNTVNDQLDKONINLALTHY 264  
 DB 227 ---YVEBAGSGALIRGSSRGIPDPK---HKYHNFQKIAVOINDHRIQ---PSF 274  
 QY 265 NDKQDPTDY-----APDYGNRL-AVLPGF-KPSINAIKGSLSEQPTTK- 306  
 DB 275 NGQGHNTYITESSYNLTASWREADDVNRNRNANLFEYTPDSNWTLSLKADPDYQTTKY 334  
 QY 307 -----STFNIVYHDDLMGNTINVTAYTRREKGRPYPP 339  
 DB 335 AAVNNKSGPPTDYSTWTRVYNQDL-----ENIYVRSMOTREKRP 374

RESULT 39

AAIY34512  
 ID AAIY34512 standard; Protein; 848 AA.

XX AAIY34512;

DT 25-AUG-1999 (first entry)

XX Porphyromonas gingivalis protein PG50.

KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;  
 KM vaccine; antigenic.

OS Porphyromonas gingivalis.

PN WO929870-A1.

PD 17-JUN-1999.

PF 10-DEC-1998; 98WO-AU01023.

PR 04-AUG-1998; 98AU-0005028.

PR 10-DEC-1997; 97AU-0000839.

PR 31-DEC-1997; 97AU-0001182.

PR 30-JAN-1998; 98AU-0001546.

PR 10-MAR-1998; 98AU-0002264.

PR 09-APR-1998; 98AU-0002911.

PR 23-APR-1998; 98AU-0003128.

PR 05-MAY-1998; 98AU-0003338.

PR 22-MAY-1998; 98AU-0003654.

PR 29-JUL-1998; 98AU-0004917.

XX (CSLC-) CSL LTD.

```

XX PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
XX PI Ross BC, Rothel LJ, Webb EA;
XX PI
XX DR WPI; 1999-385613/32.
XX DR N-PSDB; AAX91730.
XX PT Antigenic Porphyromonas gingivalis peptides for preventing
XX PT gingivitis
XX PS Claim 1; Page 495-496; 589pp; English.
XX PA
XX CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
XX CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
XX CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
XX CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
XX CC activity with a vaccine mechanism of action. The PG polypeptides can be
XX CC used as vaccines especially against Porphyromonas gingivalis. Probes can
XX CC be used to detect Porphyromonas gingivalis in standard hybridisation
XX CC assays. Porphyromonas gingivalis is involved in periodontal disease
XX CC especially gingivitis.
XX SQ Sequence 848 AA;

Query Match 6.2%; Score 127; DB 20; Length 848;
Best Local Similarity 20.2%; Pred. No. 0.1;
Matches 68; Conservative 55; Mismatches 117; Indels 96; Gaps 15;

QY 34 VELEPVTIDKSGMALNR---ITOMPHTTKVIYEEIQEQATGSRQLADVMAQLIPSL 90
DB 115 INLDEVIS-----ANRELTTLRLAPTLVNVLNKVFQSVNASNLAQGLSFO--PGV 164

QY 91 GVSSTGTS-NFGQT---MHGRQVQLNGVPLTGSRDISRQLNSINPNQVARI-EVL-SG 144
DB 165 RVENNCQCGFNQVRINGLDGRYAQILIDSRPIMSALAGVYGLEQIPANMIERVEVRGG 224

QY 145 ATSYGSGATGGLINIVTKSDLEEEQFETRIGVHGSKLSSEGIQVQGVAGVSENGV 204
DB 225 GSALYGSSAIAAGVNIITKEP-----SHNSFTFNESLSF---TGFSKLDNNTNF 270

QY 205 LARLDVDYRTTGA-----FDANGKRIAPPAOTDKODSKLSVNTNVDWQLDDK 254
DB 271 NASIVSDNDRAGAMVFGQARYNRHNDANDGY-----SELGKIDARSLGAHS----- 317

QY 255 QNINLALTHYNDKQDTPADYDGNRLAVLFGEKPSLNAIKG-----LSLSEQPK 303
DB 318 ---YLRLSDYSK-----LTGEFHTISEFRGGDRIDLPPHVGVAEQTD 358

QY 304 TTKSTFNINY-----HHDDLW--GNTINTNAYY 329
DB 359 HSFVSGNLKYLDFSSNYKHFFQAYTSGQIVNRKSY 394

RESULT 40.
AAY34385
ID AAY34385 standard; Protein; 878 AA.
XX AC AAY34385;
XX DT 25-AUG-1999 (first entry)
XX DE Porphyromonas gingivalis protein PG50.
XX KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
XX KW vaccine; antigenic.
XX OS Porphyromonas gingivalis.
XX PN WO9929870-A1.
XX PD 17-JUN-1999.
XX PF 10-DEC-1998; 98WO-AU01023.

```

```

XX PR 04-AUG-1998; 98AU-0005028.
XX PR 10-DEC-1997; 97AU-0000839.
XX PR 31-DEC-1997; 97AU-0001182.
XX PR 30-JAN-1998; 98AU-0001546.
XX PR 10-MAR-1998; 98AU-0002264.
XX PR 09-APR-1998; 98AU-0002911.
XX PR 23-APR-1998; 98AU-0003128.
XX PR 05-MAY-1998; 98AU-0003338.
XX PR 22-MAY-1998; 98AU-0003654.
XX PR 29-JUL-1998; 98AU-0004917.
XX PS (CSLC-) CSL LTD.
XX PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
XX PI Ross BC, Rothel LJ, Webb EA;
XX PI
XX DR WPI; 1999-385613/32.
XX DR N-PSDB; AAX91603.
XX PT Antigenic Porphyromonas gingivalis peptides for preventing
XX PT gingivitis
XX PS Claim 1; Page 351-352; 588pp; English.
XX CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
XX CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
XX CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
XX CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
XX CC activity with a vaccine mechanism of action. The PG polypeptides can be
XX CC used as vaccines especially against Porphyromonas gingivalis. Probes can
XX CC be used to detect Porphyromonas gingivalis in standard hybridisation
XX CC assays. Porphyromonas gingivalis is involved in periodontal disease
XX CC especially gingivitis.
XX SQ Sequence 878 AA;

Query Match 6.2%; Score 127; DB 20; Length 878;
Best Local Similarity 20.2%; Pred. No. 0.1;
Matches 68; Conservative 55; Mismatches 117; Indels 96; Gaps 15;

QY 34 VELEPVTIDKSGMALNR---ITOMPHTTKVIYEEIQEQATGSRQLADVMAQLIPSL 90
DB 145 INLDEVIS-----ANRELTTLRLAPTLVNVLNKVFQSVNASNLAQGLSFO--PGV 194

QY 91 GVSSTGTS-NFGQT---MHGRQVQLNGVPLTGSRDISRQLNSINPNQVARI-EVL-SG 144
DB 195 RVENNCQCGFNQVRINGLDGRYAQILIDSRPIMSALAGVYGLEQIPANMIERVEVRGG 254

QY 145 ATSYGSGATGGLINIVTKSDLEEEQFETRIGVHGSKLSSEGIQVQGVAGVSENGV 204
DB 255 GSALYGSSAIAAGVNIITKEP-----SHNSFTFNESLSF---TGFSKLDNNTNF 300

QY 205 LARLDVDYRTTGA-----FDANGKRIAPPAOTDKODSKLSVNTNVDWQLDDK 254
DB 301 NASIVSDNDRAGAMVFGQARYNRHNDANDGY-----SELGKIDARSLGAHS----- 347

QY 255 QNINLALTHYNDKQDTPADYDGNRLAVLFGEKPSLNAIKG-----LSLSEQPK 303
DB 348 ---YLRLSDYSK-----LTGEFHTISEFRGGDRIDLPPHVGVAEQTD 388

QY 304 TTKSTFNINY-----HHDDLW--GNTINTNAYY 329
DB 389 HSFVSGNLKYLDFSSNYKHFFQAYTSGQIVNRKSY 424

RESULT 41
ABB52577
ID ABB52577 standard; Protein; 654 AA.
XX AC ABB52577;
XX DT 11-FEB-2002 (first entry)

```



XX Escherichia coli polypeptide SEQ ID NO 527.  
 DE  
 XX Escherichia coli; B2/D+A-; anti-inflammatory; antibacterial;  
 XX immunosuppressive; extra-intestinal infection; phylogeny; meningitis;  
 KM systemic infection; non-diarrhoeal infection; septicæmia;  
 KM pyelonephritis; antibiotic resistance.  
 XX Escherichia coli.  
 OS  
 XX MO200166572-A2.  
 PN  
 XX 13-SEP-2001.  
 PD  
 XX 12-MAR-2001; 2001MO-EP03445.  
 PF  
 XX 10-MAR-2000; 2000FR-0003145.  
 PR  
 XX 02-FEB-2001; 2001FR-0001449.  
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
 PI Bingen E, Bonaccorsi S, Clermont O, Nassif X, Tinsley C;  
 XX MPI; 2001-550253/61.  
 DR  
 XX A library of DNA fragments of Escherichia coli strains for the  
 PT phylogenetic determination of a given strain comprises polynucleotides of  
 PT nature B2/D+ A- -  
 XX  
 XX Example 6; Fig 6; 646pp; English.  
 PS  
 XX The invention relates to a library of DNA fragments of Escherichia coli  
 CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)  
 CC and encoded proteins (AB852459-AB852919 and AB852954-AB853094) of nature  
 CC B2/D+A-. The polynucleotides have potential anti-inflammatory,  
 CC antibacterial and immunosuppressive activity as part of pharmaceutical  
 CC compositions used to treat, palliate or prevent extra-intestinal E. coli  
 CC infections. The polypeptides are useful for determining the phylogenetic  
 CC group of a given E. coli strain. These polypeptides can detect and treat  
 CC an undesired development of E. coli, particularly an extra-intestinal  
 CC infection that include systemic and non-diarrhoeal infections such as  
 CC septicæmia, pyelonephritis and meningitis this is particularly  
 CC advantageous as bacterial resistance is increasing with the more  
 CC frequent use of broad spectrum antibiotics.  
 CC  
 XX  
 XX Sequence 654 AA;  
 SQ  
 Query Match 6.2%; Score 126; DB 22; Length 654;  
 Best Local Similarity 22.1%; Pred. No. 0.084;  
 Matches 103; Conservative 55; Mismatches 163; Indels 146; Gaps 26;

DB 299 --TTITNGHNKTAERNRNRTSGVNVNRSHTLIFGAHQLSGAEVYRQOKPEGSATL 356  
 QY 337 YP-----FVAPPSIAKALPI-----LQSMULPSATLDAYTKAPOA----- 371  
 DB 357 YPESGNIIDFISLRQDMTKSYFVNIIIVGSRIDRYKSPFRACELKAERLSPPAAISVSP 416  
 QY 372 ---RAYGVLOSSEKAEVLGR-----VPMIN-KPK 396  
 DB 417 TDWLMVYGSISSAFRAPVTAEMVYRDDVHFYRKQKPNVYVNLNKPKE 463  
 RESULT 42  
 AAY38832  
 ID AAY38832 standard; Protein; 725 AA.  
 XX  
 AC AAY38832;  
 DT 08-OCT-1999 (first entry)  
 XX  
 DE Neisseria meningitidis strain A antigen encoded by ORF23.  
 XX  
 KM Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KM treatment; Neisseria infection; meningitis; septicæmia; gonorrhea.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN MO9924578-A2.  
 XX  
 PD 20-MAY-1999.  
 XX  
 PF 09-OCT-1998; 98WO-IB01665.  
 XX  
 PR 01-SEP-1998; 98GB-0019016.  
 PR 06-NOV-1997; 97GB-0023516.  
 PR 14-NOV-1997; 97GB-0024190.  
 PR 18-NOV-1997; 97GB-0024386.  
 PR 27-NOV-1997; 97GB-0025158.  
 PR 10-DEC-1997; 97GB-0026147.  
 PR 14-JAN-1998; 98GB-0000759.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 XX  
 PI Grandi G, Maignani V, Pizsa M, Rappuoli R, Scarlato V;  
 XX  
 DR MPI; 1999-327407/27.  
 XX  
 DR N-PSDB; AA212260.  
 XX  
 XX Proteins from Neisseria meningitidis and N. gonorrhoeae useful for  
 PT diagnosis, treatment and prevention of infection  
 PT  
 PS Claim 4; Page 381; 524pp; English.  
 XX  
 XX Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis  
 CC and N. gonorrhoeae antigenic proteins. They are encoded by open  
 CC reading frames (ORFs) AA21972-21358. The antigenic proteins,  
 CC their fragments, their nucleic acids and antibodies are used for  
 CC diagnosis, prevention (as vaccines) or treatment of Neisseria  
 CC infections, such as meningitis, septicæmia and gonorrhea. Both  
 CC organisms are closely related. Fragments of the nucleic acids  
 CC are useful as hybridisation probes and antisense reagents.  
 CC  
 XX  
 XX Sequence 725 AA;  
 SQ  
 Query Match 6.1%; Score 125.5; DB 20; Length 725;  
 Best Local Similarity 19.6%; Pred. No. 0.11;  
 Matches 74; Conservative 62; Mismatches 155; Indels 87; Gaps 16;

Db 68 LREIPQSVSVITSQMRQONIKALDRALLOQTGTSRQIYG-----SDRAGYNVL 116  
 Qy 102 QTMHGRQVQFLNGVPLTGSRDISRQLNSINPNQVARIEVLGATSIY-GSGATGGLINI 160  
 Db 117 FARGSRIRANYQINGIPVA--DALADTGNANTAAVERVEVVRGVAGLLDGTGEPSTVNL 173  
 Qy 161 VTKSLDEEQPETRIGVHGKSLSSGIGYQVQCSVAG-VSENGVNLARLDVDRYTTGGAF 219  
 Db 174 VRKRPTRKPLFEVRAEAGNRK-----HFGLGADVSGSLNAEGTLRGRL-----VSTFGRG 223  
 Qy 220 DANGKRIAPEPAQTDKQSKLSLVNTVDWDLDDKQNLNLTALTHYNDKQDTD-----YAPD 275  
 Db 224 DSWROR-----EKSRLDELVCILEYDIAPQTRVHAGMDYQOAKETADAPLSYAVY 273  
 Qy 276 YGNRLAVLFGKE--PSLNAIKGLSLEQPKTKSTFNINYYHHDDLWGNTINTNAYRREK 333  
 Db 274 DSQGVATAFGPKDNPATN-----WANSRHRALNLPAGIEHRENDQW----KLKAEYDVTR 324  
 Qy 334 GRFYPFVAPFSAKALPI 351  
 Db 325 SRFR---QPYGVAGVLSI 339

RESULT 43  
 ID: AAW01462 standard; Protein; 719 AA.  
 XX AAW01462;  
 AC AAW01462;  
 XX 24-FEB-1997 (first entry)  
 DT NTHI HxuC protein.  
 DE HxuC; NTHI; vaccine; genetic immunisation; diagnosis;  
 KW meningitis; pneumonia; bacteraemia; otitis media.  
 XX Haemophilus influenzae nontypeable strain N182.  
 OS  
 FH Key Location/Qualifiers  
 FT Region 6..18  
 FT /label= TonB box  
 FT Cleavage-site 19..21  
 FT /note= "signal peptidase I cleavage site"  
 XX WO9633275-A1.  
 PN 24-OCT-1996.  
 PD 15-APR-1996; 96WO-US05167.  
 XX 20-APR-1995; 95US-0425843.  
 XX (TEXA ) UNIV TEXAS SYSTEM.  
 PA Cope LD, Hansen EJ, Hanson MS, Jarosik GP;  
 PI WPI; 1996-485781/48.  
 XX N-PSDB; AAT44520.  
 DR Genes encoding H. influenzae HxuC and HxuD surface-expressed  
 PT protein(s) - useful in the prepn. of vaccines for children against  
 PT H. influenzae infection  
 XX Claim 11; Page 153-159; 180pp; English.  
 PS The HxuC protein (AAW01462, see also AAW01461) of nontypeable  
 CC Haemophilus influenzae (NTHI) strain N182 is a 78 kDa outer  
 CC membrane protein involved in the utilisation of low levels of free  
 CC haem. It shows homology to TonB-dependent outer membrane proteins  
 CC of other bacteria, and was identified from an open reading frame  
 CC (see also AAT44520) isolated by PCR amplification. Another outer  
 CC membrane protein, HxuD (AAW01465, see also AAW01464), was similarly  
 CC identified. Recombinant HxuC and HxuD proteins can be prepd. in

CC transformed host cells and used: to prepare vaccines against NTHI  
 CC infection; to raise diagnostic antibodies; and to prepare NTHI  
 CC diagnostic or therapeutic compans.  
 XX SQ Sequence 719 AA;  
 Query Match 6.1%; Score 125; DB 17; Length 719;  
 Best Local Similarity 23.8%; Pred. No. 0.12;  
 Matches 86; Conservative 52; Mismatches 135; Indels 88; Gaps 19;  
 Qy 7 FQWLSLPLLSVAVTQQLVAQPNESLPTVELEP--VWITIDKSGMALANRITOMPHTTKVI 64  
 Db 3 FSKLSLAITTLVLTANALAQ-----SVELDSINVIAIATRPSRFAYT----- 43  
 Qy 65 YEEQIQEQATGSRQLADVMAQL--IPSLGVSSGTTS-----NFGQTMHGRQVQFLNGVP 117  
 Db 44 -PEKQSKDLSLKQATSVAAALEDPNVVVRGGSRSIAQKPNIRGLSDNRVQV-VIDGV- 100  
 Qy 118 LTGSRDISRQLNSINP-NOVARIEVLG-ATSIYGSGATGGLINIVTKSLDEEQFETRI 175  
 Db 101 -RQNFDLAHRGSYFLPMSLIQIEIEVIKGPSSSLWGSALGGVAVARTPNALDLKNNDKF 159  
 Qy 176 GVHGSKLSSEGIGYQVQGVAGVSENGVNLARLDVDRYTTGGAFDA--NGKRIAPEPAQ 233  
 Db 160 GV---KIRQ---GYQ-----TANNLSERDVSVFAANDKFDVLISGFYNNADNLT 203  
 Qy 234 DKQDS-----KSLSVNTNVDWDLDDKQNLNLA-----LTHYNDKQDPTD 271  
 Db 204 GRGNKLNNTAVKQFGLAKFGWSINDANRVELSHRETRFKQTAPSNNEVENELTNKQITD 263  
 Qy 272 YAPDYGNRLAVLFGKEKPSLNAI-KGLS-----LSEQPKTKST-FNINHHDDLWGN 321  
 Db 264 QIEYHRSKKT---KPSLEEFYSGVKTRFGSVSYLSDQOIPDQSTVFNNYLYLTPDNPYLN 319  
 Qy 322 T 322  
 Db 320 T 320

RESULT 44  
 AAW55247  
 ID AAW55247 standard; Protein; 529 AA.  
 XX AAW55247;  
 AC AAW55247;  
 DT 26-JUN-1998 (first entry)  
 DE H. pylori ORF 05ap21216orf4 protein.  
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 KW identification; binding compound; bacteria; life cycle; activator;  
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;  
 KW bacterium.  
 XX Helicobacter pylori.  
 OS WO9737044-A1.  
 PN 09-OCT-1997.  
 PD 27-MAR-1997; 97WO-US05223.  
 XX 06-DEC-1996; 96US-0761318.  
 PR 29-MAR-1996; 96US-0625811.  
 PR 02-APR-1996; 96US-0758731.  
 PR 25-OCT-1996; 96US-0736905.  
 PR 28-OCT-1996; 96US-0738859.  
 XX (ASTR ) ASTRA AB.  
 XX Alm RA, Smith D;  
 XX WPI; 1997-503122/46.

DR N-PSDB; AAV24656.  
 XX Helicobacter pylori nucleic acid sequences and encoded  
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
 PT infection and for diagnosis of H. pylori infection  
 XX Claim 14; Pages 489-490; 1145pp; English.  
 XX This sequence is a Helicobacter pylori protein of unspecified  
 CC function.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors.  
 CC The DNA and probes derived from it may be used for the  
 CC identification of H. pylori in a sample, and the diagnosis of  
 CC H. pylori infection. Nucleic acid sequences complementary to the  
 CC DNA act as antisense sequences, and can be used to prevent the  
 CC translation of H. pylori mRNA. Antibodies against the protein can  
 CC be used in immunoassays to evaluate the abundance and distribution  
 CC of H. pylori-specific antigens. The genomic sequence of H. pylori  
 CC (ATCC 55679) was determined from overlapping contigs generated by  
 CC mechanically shearing the bacterial DNA. The sequences were  
 CC analysed for ORF of at least 180 nucleotides, and the predicted  
 CC coding regions defined by computer evaluation, to identify likely  
 CC H. pylori antigens for vaccine development, the amino acid  
 CC sequences predicted from various ORF were analysed for significant  
 CC homology to other known or exported membrane proteins. Having  
 CC identified and determined the sequences of interest, particular  
 CC regions can be isolated from H. pylori by PCR amplification for  
 CC recombinant polypeptide production, e.g. in E. coli hosts.  
 XX  
 XX Sequence 529 AA;  
 SQ  
 Query Match 6.0%; Score 123.5; DB 18; Length 529;  
 Best Local Similarity 22.7%; Pred. No. 0.1; Indels 89; Gaps 22;  
 Matches 90; Conservative 48; Mismatches 169; Indels 89; Gaps 22;  
 QY 2 RHSHYFQWMLSLPLSLVAVVOQLYAOQNESLPVLEBPVITIDKSGMALNRITQMPHTT 61  
 DB 36 KDKHNF-----LKKVTTTEQKF---SSSAPISWSEBVNNTSSRTVISNK--ELKKTG 84  
 QY 62 KVIYEEQIQ-----EQATGSRQADVMAQLIPSLVSSGTTSNFGQTMHGRQVOFLN 114  
 DB 85 NLIENALQWVPGIQRDADATGTVLPKISYR-----GFGGGGNG-----HSNTMTLVN 133  
 QY 115 GVPLTGS--RDISRQNSINPNQVARIIEVLSGATSI--YSGSGATGSLINITYKSDLF--EE 169  
 DB 134 GIPIYGAPYSNIELAIPVTFQSVDRIDVILKGGTSVOYGPNTFGGVNVTIKETPKEMEN 193  
 QY 170 QPETRLGVHSGKSLSESGIGYOVQSVAGVSENGNVLAR-----LDVDYRTTGG----- 217  
 DB 194 QAABERITFWER--SSNG-----NFPDPEKKEKPLAQTLGNQMLFNTYGRITAGMLCKYI 244  
 QY 218 AFDANGKRIAPBPQTDKQDSKSLSVTNVDMQJDKONINLALT--HYNDKQDTPDYAP 274  
 DB 245 GISAOGNMI---NOQGFROMSPTKVN---YLLDAIKYNATYTFKAYQYQYNSYHP 297  
 QY 275 -----DYG--NRLAVLFGKPSNLAIKGLS-----LSBPKTKTKSTFNINHYHDD 317  
 DB 298 GTTSAQOYAYNR--FINEBPD--NODGGRARFGIYQNTFGDPRKVGSDFKFTYFTHD 353  
 QY 318 L---WGNTINTNAYYREKGRFPYFVAPFSIAKALP 350  
 DB 354 MSRDGFSNQYQSVYVMSGCKILPFRKKGKGEISAKNP 389

XX DE H. pylori ORF 07ap20216\_7227202\_f3\_10 cell envelope OMP.  
 XX KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;  
 XX KW identification; binding compound; bacteria; life cycle; activator;  
 XX KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.  
 XX Helicobacter pylori.  
 XX  
 XX WO9737044-A1.  
 XX  
 XX 09-OCT-1997.  
 XX  
 XX 27-MAR-1997; 97WO-US05223.  
 XX  
 XX 06-DEC-1996; 96US-0761318.  
 XX 29-MAR-1996; 96US-0625811.  
 XX 02-APR-1996; 96US-0758731.  
 XX 25-OCT-1996; 96US-0736905.  
 XX 28-OCT-1996; 96US-0738859.  
 XX  
 XX (ASTR ) ASTRA AB.  
 XX  
 XX Alm RA, Smith D;  
 XX  
 XX WPI; 1997-503122/46.  
 XX  
 XX N-PSDB; AAV24685.  
 XX  
 XX Helicobacter pylori nucleic acid sequences and encoded  
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori  
 PT infection and for diagnosis of H. pylori infection  
 XX  
 XX Claims 14,80; Page 683-684; 1145pp; English.  
 PS  
 XX This sequence is stated as being a H. pylori cell envelope outer membrane  
 CC protein (OMP) having a terminal phe residue.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors. The  
 CC DNA and probes derived from it may be used for the identification of  
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic  
 CC acid sequences complementary to the DNA act as antisense sequences and  
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies  
 CC against the protein can be used in immunoassays to evaluate the abundance  
 CC and distribution of H. pylori-specific antigens. The genomic sequence of  
 CC H. pylori (ATCC 55679) was determined from overlapping contigs generated  
 CC by mechanically shearing the bacterial DNA. The sequences were analysed  
 CC for ORF of at least 180 nucleotides, and the predicted coding regions  
 CC defined by computer evaluation. To identify likely H. pylori antigens for  
 CC vaccine development, the amino acid sequences predicted from various ORF  
 CC were analysed for significant homology to other known or exported  
 CC membrane proteins. Having identified and determined the sequences of  
 CC interest, particular regions can be isolated from H. pylori by PCR  
 CC amplification for recombinant polypeptide production, e.g. in E. coli  
 CC hosts.  
 CC  
 CC Sequence 533 AA;  
 SQ  
 Query Match 6.0%; Score 123.5; DB 18; Length 533;  
 Best Local Similarity 22.7%; Pred. No. 0.1;  
 Matches 90; Conservative 48; Mismatches 169; Indels 89; Gaps 22;  
 QY 2 RHSHYFQWMLSLPLSLVAVVOQLYAOQNESLPVLEBPVITIDKSGMALNRITQMPHTT 61  
 DB 36 KDKHNF-----LKKVTTTEQKF---SSSAPISWSEBVNNTSSRTVISNK--ELKKTG 84  
 QY 62 KVIYEEQIQ-----EQATGSRQADVMAQLIPSLVSSGTTSNFGQTMHGRQVOFLN 114  
 DB 85 NLIENALQWVPGIQRDADATGTVLPKISYR-----GFGGGGNG-----HSNTMTLVN 133  
 QY 115 GVPLTGS--RDISRQNSINPNQVARIIEVLSGATSI--YSGSGATGSLINITYKSDLF--EE 169  
 DB 134 GIPIYGAPYSNIELAIPVTFQSVDRIDVILKGGTSVOYGPNTFGGVNVTIKETPKEMEN 193



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: December 25, 2002, 20:18:38 ; Search time 25 Seconds  
(without alignments)  
470.767 Million cell updates/sec

Title: US-09-889-746-2

Perfect score: 2047  
Sequence: 1 MEHSHYFQWLSPLSLVAVT.....SKAEVLGRVPIKPKRALF 400

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A COMB pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B COMB pep:\*  
3: /cgn2\_6/prodata/1/1aa/5A COMB pep:\*  
4: /cgn2\_6/prodata/1/1aa/5B COMB pep:\*  
5: /cgn2\_6/prodata/1/1aa/PCTUS COMB pep:\*  
6: /cgn2\_6/prodata/1/1aa/backfillseq pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	191.5	9.4	725	4	US-09-668-113A-8 Sequence 8, App1
2	182.5	8.9	117	1	US-08-249-013-6 Sequence 6, App1
3	182.5	8.9	117	2	US-08-886-863-6 Sequence 6, App1
4	182.5	8.9	117	4	US-09-175-229-6 Sequence 6, App1
5	182.5	8.9	117	5	PCT-US95-06764-6 Sequence 6, App1
6	147.5	7.2	970	1	US-08-375-709-7 Sequence 7, App1
7	147.5	7.2	970	1	US-08-752-929-7 Sequence 7, App1
8	147.5	7.2	970	4	US-09-090-793-5 Sequence 5, App1
9	141	6.9	663	3	US-08-765-081-7 Sequence 7, App1
10	141	6.9	663	3	US-09-098-082-7 Sequence 7, App1
11	137.5	6.7	652	1	US-08-765-081-6 Sequence 6, App1
12	137.5	6.7	652	3	US-09-098-082-6 Sequence 6, App1
13	136.5	6.7	1076	2	US-08-867-941-19 Sequence 19, App1
14	136.5	6.7	1076	4	US-09-074-658-19 Sequence 19, App1
15	136	6.6	696	1	US-08-765-081-5 Sequence 5, App1
16	136	6.6	696	1	US-09-098-082-5 Sequence 5, App1
17	136	6.6	696	5	PCT-US95-06994-5 Sequence 5, App1
18	135.5	6.6	703	5	PCT-US95-06994-8 Sequence 8, App1
19	134.5	6.6	718	8	PCT-US95-06994-6 Sequence 8, App1
20	134	6.5	1053	3	US-08-613-009A-8 Sequence 8, App1
21	134	6.5	1053	4	US-08-778-570B-10 Sequence 10, App1
22	134	6.5	1053	4	US-09-059-584-10 Sequence 10, App1
23	134	6.5	1074	3	US-08-613-009A-7 Sequence 7, App1
24	134	6.5	1074	4	US-08-778-570B-9 Sequence 9, App1
25	134	6.5	1074	4	US-09-059-584-9 Sequence 9, App1
26	133	6.5	715	3	US-08-425-843-7 Sequence 7, App1
27	130	6.4	1052	4	US-08-778-570B-14 Sequence 14, App1

28	130	6.4	1052	4	US-09-059-584-14 Sequence 14, App1
29	130	6.4	1070	3	US-08-613-009A-11 Sequence 11, App1
30	130	6.4	1070	4	US-08-778-570B-13 Sequence 13, App1
31	130	6.4	1070	4	US-09-059-584-13 Sequence 13, App1
32	128	6.3	790	3	US-08-537-361B-8 Sequence 8, App1
33	128	6.3	790	4	US-08-817-707-8 Sequence 8, App1
34	127	6.2	944	2	US-08-867-941-23 Sequence 23, App1
35	127	6.2	944	4	US-09-074-658-23 Sequence 23, App1
36	124.5	6.1	718	5	PCT-US95-06994-7 Sequence 7, App1
37	123.5	6.0	725	3	US-08-425-843-2 Sequence 2, App1
38	122	6.0	791	3	US-08-537-361B-4 Sequence 4, App1
39	122	6.0	791	4	US-08-817-707-4 Sequence 4, App1
40	121	5.9	790	4	US-08-817-707-6 Sequence 6, App1
41	121	5.9	909	2	US-08-363-124A-4 Sequence 4, App1
42	118.5	5.8	941	4	US-09-074-658-75 Sequence 75, App1
43	118	5.8	908	1	US-08-487-890A-94 Sequence 94, App1
44	118	5.8	908	2	US-08-478-435-94 Sequence 94, App1
45	118	5.8	908	2	US-08-337-483-94 Sequence 94, App1

ALIGNMENTS

RESULT 1  
US-09-668-113A-8  
; Sequence 8, Application US/09668113A  
; Patent No. 6410703  
; GENERAL INFORMATION:  
; APPLICANT: Russo, Thomas A.  
; TITLE OF INVENTION: Identification of A Vaccine Candidate from an  
; FILE REFERENCE: 11520.0214  
; CURRENT APPLICATION NUMBER: US/09/668,113A  
; CURRENT FILING DATE: 2000-09-22  
; NUMBER OF SEQ ID NOS: 10  
; SEQ ID NO 8  
; LENGTH: 725  
; TYPE: PRT  
; ORGANISM: Baccharichia coli  
; FEATURE:  
US-09-668-113A-8

Query Match Best Local Similarity 9.4%; Score 191.5; DB 4; Length 725;  
Matches 84; Conservative 59; Mismatches 135; Indels 63; Gaps 17;

QY	9	WLSPLSLVAVTQQL-----YAQPNESLPTVELEPVYITITDKSGMALANRITQMPHTTKYI	64
DB	8	W-SLTVLVLVGLNSQVAVAKYSDDND-----ETLVE-----ATRAEVLKQOPGVSYI	54
QY	65	YEQIOEQATGSRQLADVMAQLIPSLGV-----SSGTTSPNGQT-----MGRQVQPLNGV	116
DB	55	TSEDI-KKTPPVNDLSDIIRKM-PGVNLGNSASGTRGNRRQIDIRGKPEPNTLIDGV	112
QY	117	PLT-----GSRDISRLNS-INPNQVARIELVSG-ATSIYSGATGGLINIVTKSD	165
DB	113	PVTSRNSVYSWRCGRDTRGDTNRWVPEQEYERIEVIRGPAAARYSGAAGVNIITKRP	172
QY	166	LEEBEPEFTRIGVHGSKLSSEBGIGYQVQGSVAGVSENGVARIADVDY-RTTGAAPRANGK	224
DB	173	TNDHSGLSLTYTOPSSSEBGATRRANFSLGCLADALATRLYGNLINTKTDADSWDIN--	230
QY	225	RIAPPEAQDQDK-----QDSKLSVNTNVMDQLDDKQINIALTHYNDKQOTDVAPOGRL	280
DB	231	----SPVGTNAAGHGVRNMDINGVSWKLNQQLIDEVGV--SRQGNITAGDTQNS	284
QY	281	AVLFEKPSLAIKGLSLSEOPKTKSTFNINY--HHDDL	319
DB	285	SSAVTE-----SLAKSGKETNRLRYQVYGIHNGIM	315

RESULT 2  
US-08-249-013-6

RESULT 2  
US-08-249-013-6

; Sequence 6, Application US/08249013  
; Patent No. 5843754  
; GENERAL INFORMATION:  
; APPLICANT: Haake, David A.  
; TITLE OF INVENTION: CLONED leptospira OUTER MEMBRANE PROTEIN  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Spensley Horn Jubas & Lubitz  
; STREET: 1880 Century Park East, Suite 500  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90067  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/249,013  
; FILING DATE: 25-MAY-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Tumarkin Ph.D., Lisa A.,  
; REGISTRATION NUMBER: P-38,347  
; REFERENCE/DOCKET NUMBER: PD-3602  
; TELEPHONE: (619) 455-5100  
; TELEFAX: (619) 455-5110  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 117 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; IMMEDIATE SOURCE:  
; CLONE: Iuta  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..117  
US-08-249-013-6

Query Match 8.9%; Score 182.5; DB 1; Length 117;  
Best Local Similarity 34.8%; Pred. No. 5.4e-10;  
Matches 49; Conservative 20; Mismatches 39; Indels 33; Gaps 3;

Qy 82 VMAQLIPSLGVSSGTTNFGQTMHGRQVQLLNGVPLTGSRDISRQLNSINPNQVARIIV 141  
Db 5 VSAANRIPGLDVSSRSRTNYGMVGRPLVVLVDGVRLN-----HHIEV 47

Qy 142 LSGATSIYGSATGGLINIVTKSLDLEEEQFETRIGVHGSKLSSEGIGYQVQSVAGVSEN 201  
Db 48 IFGATSLYGGSGTGLINIVTK-----LEGVKVDSYELGWRDKRRIYGVLS 94

Qy 202 GNVLARLDVDTTGGAFDAN 222  
Db 95 FSIENLFDRIYR---GRFGLN 112

RESULT 3  
US-08-886-863-6  
; Sequence 6, Application US/0886863  
; Patent No. 5824321  
; GENERAL INFORMATION:  
; APPLICANT: Haake, David A.  
; TITLE OF INVENTION: CLONED leptospira OUTER MEMBRANE PROTEIN  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Spensley Horn Jubas & Lubitz  
; STREET: 1880 Century Park East, Suite 500  
; CITY: Los Angeles  
; STATE: California

; COUNTRY: USA  
; ZIP: 90067  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/886,863  
; FILING DATE: 01-JUL-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/249,013  
; FILING DATE: 25-MAY-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Tumarkin Ph.D., Lisa A.,  
; REGISTRATION NUMBER: P-38,347  
; REFERENCE/DOCKET NUMBER: PD-3602  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 455-5100  
; TELEFAX: (619) 455-5110  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 117 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; IMMEDIATE SOURCE:  
; CLONE: Iuta  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..117  
US-08-886-863-6

Query Match 8.9%; Score 182.5; DB 2; Length 117;  
Best Local Similarity 34.8%; Pred. No. 5.4e-10;  
Matches 49; Conservative 20; Mismatches 39; Indels 33; Gaps 3;

Qy 82 VMAQLIPSLGVSSGTTNFGQTMHGRQVQLLNGVPLTGSRDISRQLNSINPNQVARIIV 141  
Db 5 VSAANRIPGLDVSSRSRTNYGMVGRPLVVLVDGVRLN-----HHIEV 47

Qy 142 LSGATSIYGSATGGLINIVTKSLDLEEEQFETRIGVHGSKLSSEGIGYQVQSVAGVSEN 201  
Db 48 IFGATSLYGGSGTGLINIVTK-----LEGVKVDSYELGWRDKRRIYGVLS 94

Qy 202 GNVLARLDVDTTGGAFDAN 222  
Db 95 FSIENLFDRIYR---GRFGLN 112

RESULT 4  
US-09-175-229-6  
; Sequence 6, Application US/09175229  
; Patent No. 6309641  
; GENERAL INFORMATION:  
; APPLICANT: Haake, David A.  
; TITLE OF INVENTION: CLONED leptospira OUTER MEMBRANE PROTEIN  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Spensley Horn Jubas & Lubitz  
; STREET: 1880 Century Park East, Suite 500  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90067  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/175,229  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/249,013  
FILING DATE: 25-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Tumarkin Ph.D., Lisa A.,  
REGISTRATION NUMBER: P-38,347  
REFERENCE/DOCKET NUMBER: PD-3602  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 455-5100  
TELEFAX: (619) 455-5110  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 117 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
IMMEDIATE SOURCE:  
CLONE: IuCa  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..117  
US-09-175-229-6

Query Match 8.9%; Score 182.5; DB 4; Length 117;  
Best Local Similarity 34.8%; Pred. No. 5.4e-10;  
Matches 49; Conservative 20; Mismatches 39; Indels 33; Gaps 3;

QY 82 VMAQLPSLGVSSTGTFNFGQTMHGRVQFLNGVPLTGSRDISRQLNSINPQVARIIV 141  
DB 5 VSANRIPGLDVSSRSRTNYGMVGRPLVLDGVRNL-----HHIEV 47  
QY 142 LSGATSIYSGATGGLINITYKSDLSEEQFETRIGVHSGKLSSEGIQYGVGQSVAGVSEN 201  
DB 48 IFGATSLYGGSGTGLINITYK-----LEGVKVDSYELGMDKRRIRYGVELS 94  
QY 202 GNVTLARLDVDRRTTGAFDPAN 222  
DB 95 FSIENLPDRDYR--GRFGLN 112

RESULT 5  
PCT-US95-06764-6  
Sequence 6, Application PC/TUS9506764  
GENERAL INFORMATION:  
APPLICANT: The Regents of the University of California  
TITLE OF INVENTION: CLONED leptospira OUTER MEMBRANE PROTEIN  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: California  
COUNTRY: USA  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/06764  
FILING DATE: 25-MAY-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Halle, Ph.D., Lisa A.,  
REGISTRATION NUMBER: 38,347,  
REFERENCE/DOCKET NUMBER: FD3602  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 678-5070

TELEFAX: (619) 678-5099  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 117 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
IMMEDIATE SOURCE:  
CLONE: IuCa  
FEATURE:  
NAME/KEY: Protein  
LOCATION: 1..117  
PCT-US95-06764-6

Query Match 8.9%; Score 182.5; DB 5; Length 117;  
Best Local Similarity 34.8%; Pred. No. 5.4e-10;  
Matches 49; Conservative 20; Mismatches 39; Indels 33; Gaps 3;

QY 82 VMAQLPSLGVSSTGTFNFGQTMHGRVQFLNGVPLTGSRDISRQLNSINPQVARIIV 141  
DB 5 VSANRIPGLDVSSRSRTNYGMVGRPLVLDGVRNL-----HHIEV 47  
QY 142 LSGATSIYSGATGGLINITYKSDLSEEQFETRIGVHSGKLSSEGIQYGVGQSVAGVSEN 201  
DB 48 IFGATSLYGGSGTGLINITYK-----LEGVKVDSYELGMDKRRIRYGVELS 94  
QY 202 GNVTLARLDVDRRTTGAFDPAN 222  
DB 95 FSIENLPDRDYR--GRFGLN 112

RESULT 6  
US-08-375-709-7  
Sequence 7, Application US/08375709  
Patent No. 5683898  
GENERAL INFORMATION:  
APPLICANT: YAZAWA, Kazunaga  
APPLICANT: YAMADA, Akiko  
APPLICANT: KATO, Seishi  
TITLE OF INVENTION: Gene Coding For Eicosapentaenoic Acid  
TITLE OF INVENTION: Synthesizing Enzymes and Process for Production of  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/375,709  
FILING DATE: 20-JAN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/178,251  
FILING DATE: 14-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-147945  
FILING DATE: 15-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: WEGNER, Harold C.,  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 53466/150/AAOK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300





RESULT 9  
US-08-765-081-7  
Sequence 7, Application US/08765081  
Patent No. 5786260  
GENERAL INFORMATION:  
APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.  
TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Christensen, O'Connor, Johnson and Kindness  
STREET: 2800 Pacific First Center, 1420 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98101-2347  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette-3.5 inch, 1.44MB storage  
COMPUTER: IBM PC compatible/Pentium  
OPERATING SYSTEM: MS-Windows 3.1  
SOFTWARE: Word for Windows-6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/765,081  
FILING DATE: March 26, 1997  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/06994  
FILING DATE: June 7, 1995  
APPLICATION NUMBER: US 08/265,714  
FILING DATE: June 24, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Shelton, Dennis K.  
REGISTRATION NUMBER: 26,997  
REFERENCE/DOCKET NUMBER: CHOR-1-10286  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)  
TELEFAX: 1-206-224-0779  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 663 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: E. coli C1ra protein amino acid sequence  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Escherichia Coli  
US-08-765-081-7  
Query Match 6.9%; Score 141; DB 1; Length 663;  
Best local Similarity 21.2%; Pred. No. 8.5e-05;  
Matches 65; Conservative 53; Mismatches 99; Indels 90; Gaps 12;

QY 281 AVLGEK 287  
DB 292 LKRYGEK 298  
RESULT 10  
US-09-098-082-7  
Sequence 7, Application US/09098082  
Patent No. 6040421  
GENERAL INFORMATION:  
APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.  
TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Christensen, O'Connor, Johnson and Kindness  
STREET: 2800 Pacific First Center, 1420 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98101-2347  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette-3.5 inch, 1.44MB storage  
COMPUTER: IBM PC compatible/Pentium II  
OPERATING SYSTEM: MS-Windows 95  
SOFTWARE: Word for Windows-6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/098,082  
FILING DATE:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/765,081  
FILING DATE: March 26, 1997  
APPLICATION NUMBER: PCT/US95/06994  
FILING DATE: June 7, 1995  
APPLICATION NUMBER: US 08/265,714  
FILING DATE: June 24, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Shelness, Diana K.  
REGISTRATION NUMBER: 35,356  
REFERENCE/DOCKET NUMBER: CHOR-1-12402  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)  
TELEFAX: 1-206-224-0779  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 663 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: E. coli C1ra protein amino acid sequence  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Escherichia Coli  
US-09-098-082-7  
Query Match 6.9%; Score 141; DB 3; Length 663;  
Best local Similarity 21.2%; Pred. No. 8.5e-05;  
Matches 65; Conservative 53; Mismatches 99; Indels 90; Gaps 12;

Qy 242 --SVNTNVDWLDKQNTLNALTHYNDKQDTP-----YAP-----DYGNRL 280  
Db 232 FSSRDGNVEFAWTPNQNHDTAGYGFDRQDRDSDLDKNRLERQNSVSHNGRWDTGTSE 291  
Qy 281 AVLFGK 287  
Db 292 LKYYGK 298  
RESULT 11  
US-08-765-081-6  
; Sequence 6, Application US/08765081  
; Patent No. 5798260  
; GENERAL INFORMATION:  
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.  
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness  
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98101-2347  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage  
; COMPUTER: IBM PC compatible/Pentium  
; OPERATING SYSTEM: MS-Windows 3.1  
; SOFTWARE: Word for Windows-6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/765.081  
; FILING DATE: March 26, 1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/06994  
; FILING DATE: June 7, 1995  
; APPLICATION NUMBER: US 08/265.714  
; FILING DATE: June 24, 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Shelton, Dennis K.  
; REGISTRATION NUMBER: 26,997  
; REFERENCE/DOCKET NUMBER: CHOR-1-10286  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)  
; TELEFAX: 1-206-224-0779  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 652 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; DESCRIPTION: Vibrio cholerae Irga amino acid sequence  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Vibrio Cholerae  
US-08-765-081-6

Query Match 6.7%; Score 137.5; DB 1; Length 652;  
Best Local Similarity 22.6%; Pred. No. 0.00018;  
Matches 82; Conservative 47; Mismatches 115; Indels 119; Gaps 18;

Qy 37 EPVVITIDKSGMALANRITQMPHTTKVIYEIQIQEQATGSRQLADVMAQL--IPSLGVSS 94  
Db 33 ETMVVT---AAGYAQVIONAPASISVISREDLE-----SRYRDVTDALKSVPGVTVG 83  
Qy 95 GTTSNFGQT----MGRQVQFLNGVPLTGSRRDISROL--NSINP-----NQVAR 138  
Db 84 G-----GDTTDSIRGMSNYTL--ILVDGKRQTSRQTRPNSDGGFIEQGWLPPLQAIER 136  
Qy 139 IEVLGSGATS-IYGSGATGGLINIVTKSDLEEEQFETRIGVHSGKLSSESIGIQVQGVQSVAG 197  
Db 137 IEVIRGPMSTLYGSDAIGGVINIIRTKD--QQQWSGNV-----QLSTV 177

Qy 198 VSENGVILRLVDYRTTGGAFDANGKRIAPPEPAQTQKD-----SKSL-SVNTNVDWQ 250  
Db 178 VOENRASGDEOSANFFVTGFLSDALSGLQVYGTQTTQRDEDEIEHGYGDKSLRSLTSKLNQ 237  
Qy 251 LD-----DKQN-----INLALTHYNDKQDTPDY 272  
Db 238 LNPDHQLOLEAGVSAQDRNNVNGSAOSSGCRGTCSNTDQYRRNHVAVSHQGDWQGVQ 297  
Qy 273 APDYGNRILAVLFGKPSLNALKGLSLSEQPKTKST-----FNINYHDDLMGN 321  
Db 298 SDTY-----LQYEE-----NTNKSREMSIDNTVPKSTLVAPIGEHMLSGVGEKGHESLEDK 348  
Qy 322 TIN 324  
Db 349 TSN 351  
RESULT 12  
US-09-098-082-6  
; Sequence 6, Application US/09098082  
; Patent No. 6040421  
; GENERAL INFORMATION:  
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.  
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness  
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98101-2347  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage  
; COMPUTER: IBM PC compatible/Pentium II  
; OPERATING SYSTEM: MS-Windows 95  
; SOFTWARE: Word for Windows-6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/098,082  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/765,081  
; FILING DATE: March 26, 1997  
; APPLICATION NUMBER: PCT/US95/06994  
; FILING DATE: June 7, 1995  
; APPLICATION NUMBER: US 08/265,714  
; FILING DATE: June 24, 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Shelness, Diana K.  
; REGISTRATION NUMBER: 35,356  
; REFERENCE/DOCKET NUMBER: CHOR-1-12402  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)  
; TELEFAX: 1-206-224-0779  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 652 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; DESCRIPTION: Vibrio cholerae Irga amino acid sequence  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Vibrio Cholerae  
US-09-098-082-6

Query Match 6.7%; Score 137.5; DB 3; Length 652;  
Best Local Similarity 22.6%; Pred. No. 0.00018;  
Matches 82; Conservative 47; Mismatches 115; Indels 119; Gaps 18;  
Qy 37 EPVVITIDKSGMALANRITQMPHTTKVIYEIQIQEQATGSRQLADVMAQL--IPSLGVSS 94  
Db 33 ETMVVT---AAGYAQVIONAPASISVISREDLE-----SRYRDVTDALKSVPGVTVG 83

QY 95 GTTSMFGOT---MHGRQVFLNGVPLTGSRDISROL--NSINP-----NOVAR 138  
 Db 84 G-----GDTTDISIRGKSNVTL--ILVDGRQTSRQTRNSDGPGLIEQGLPPLQAIER 136  
 QY 139 IEVLISGATS-IYSGGATGGLINITYKSDLEBEOPETRIGVHGKLSSEGIYGVQGSVAG 197  
 Db 137 IEVIRGPMSTLYGSDAIGVINITRD--COQWSGNV-----QLSTV 177  
 QY 198 VSENGVNLALVDVYRTTGAFFDANGRIAPPAQTDKOD-----SKSL-SVNTVNDWQ 250  
 Db 178 VQENRASGDEQSANFPTGPIUSDALSLQVYGQTTQRDEDELEHGYSKSLRSLTSKLNQ 237  
 QY 251 LD-----DKON-----INLATHYNDKQDPTDY 272  
 Db 238 LNPDLQLEAGVSAQDRNNVVKSAQSSGCRGTCSTNDQYRNNHVAVSHQDMQGVQ 297  
 QY 273 APDYGNFLAVFGKPSLMAIKGLSLSEOPKTKST-----FNINYHDDIMGN 321  
 Db 298 SDTY-----LQYER-----NTTKSREMSIDMTVFKSTLVAPIGEMLSFGVEGKESLEDK 348  
 QY 322 TIN 324  
 Db 349 TSN 351

## RESULT 13

US-08-867-941-19  
 ; Sequence 19, Application US/08867941  
 ; Patent No. 5977337

GENERAL INFORMATION:  
 APPLICANT: Loomore, Sheena M  
 APPLICANT: Du, Run-Pan  
 APPLICANT: Wang, Quijun  
 APPLICANT: Yang, Yan-Ping  
 APPLICANT: Klein, Michel H  
 TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA  
 NUMBER OF SEQUENCES: 67  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sim & McBurney  
 STREET: 6th Floor, 330 University Avenue  
 CITY: Toronto  
 STATE: Ontario  
 COUNTRY: Canada  
 ZIP: M5G 1R7  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/867,941  
 FILING DATE: 03-JUN-1997  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Stewart, Michael I  
 REGISTRATION NUMBER: 24,973  
 TELECOMMUNICATION INFORMATION:  
 REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb  
 TELEPHONE: (416) 595-1155  
 TELEFAX: (416) 595-1163  
 INFORMATION FOR SEQ ID NO: 19:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1076 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-867-941-19

Query Match 6.7%; Score 136.5; DB 2; Length 1076;  
 Best Local Similarity 23.6%; Pred. No. 0.0005;  
 Matches 95; Conservative 48; Mismatches 150; Indels 109; Gaps 21;

QY 10 LSLPLLSVA-----VTQQLVAQPNESLPTVELEPVITIDKSG--MALANRITOMPTT 61  
 Db 22 LSLGLINITQVALANTATADKAETDKTNLVLVLEDTVTAKKNAFVSRANNEVTGIGKV 81  
 QY 62 KVIYEQIOEQATGSQLADWAQILPSLG-VSSGTTSMFGQTMHG---RQVQFLNGV- 116  
 Db 82 KVA-ETINKEQVINITRDL---TRYDPGLAVVEQGGASSGSIRKMDKRVAVLVGIN 136  
 QY 117 -----PLTGSR-DISROLINSINPQVARIIVLSGA-TSIVSGATGGLINITYKS- 164  
 Db 137 QAQHYGQPVAGKRVYAAAGALINEIYEVNVSVEISKANSSEVIGSGALSSVAFVTTAD 196  
 QY 165 ---DLEBEOPETR-----IGVHGKLSSEGI-----GYQVG 192  
 Db 197 IIKDQMGVQTKTAVASKNNAVNSVAAAGKAGSPSGLIITDRRGQREKADHDAYQCS 256  
 QY 193 QS-----VAGVSENGNV-----LALDVIDYRTTGAFFDANG-RR 225  
 Db 257 QSPDRAVATTPNNRTPFLANECANGNYEACAGGQTKLQAKPTNRDVRNVVDYGPNR 316  
 QY 226 IAPPAQTDKQDSKLSVNTNVNDWQLDKRONIN--LATHYNDKQDPTDYAPDYGNFLAVL 283  
 Db 317 LIPNPL---TQDSKSLLRPG--YQUNDKHYGVGYEITKQNYAMQDKTVPAY---LAVH 368  
 QY 284 FGKPSLN-----AIKGLSLSEOPKTK---KSTFNINYH 315  
 Db 369 DIKSRLSNHAQANGYYQGNLGERIRDTIGPDSGYGINVAH 410

## RESULT 14

US-09-074-658-19  
 ; Sequence 19, Application US/09074658  
 ; Patent No. 6184371

GENERAL INFORMATION:  
 APPLICANT: Loomore, Sheena M  
 APPLICANT: Run-Pan Du  
 APPLICANT: Quijun Wang  
 APPLICANT: Yang, Yan-Ping  
 APPLICANT: Klein, Michel H  
 TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA  
 NUMBER OF SEQUENCES: 78  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Sim & McBurney  
 STREET: 6th Floor, 330 University Avenue  
 CITY: Toronto  
 STATE: Ontario  
 COUNTRY: Canada  
 ZIP: M5G 1R7  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/074,658  
 FILING DATE: 08-MAY-1998  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Stewart, Michael I  
 REGISTRATION NUMBER: 24,973  
 TELECOMMUNICATION INFORMATION:  
 REFERENCE/DOCKET NUMBER: 1038-795  
 TELEPHONE: (416) 595-1155  
 TELEFAX: (416) 595-1163  
 INFORMATION FOR SEQ ID NO: 19:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1076 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-09-074-658-19

Query Match 6.7%; Score 136.5; DB 4; Length 1076;

Best Local Similarity 23.6%; Pred. No. 0.0005;  
Matches 95; Conservative 48; Mismatches 150; Indels 109; Gaps 21;

```

Qy 10 LSLPLLSVA-----VTOOLYAQNPSLEPTVELEPWITIDKSG--MALANRITOMPHTT 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22 LSLGLLNTVALANTTADKABATDKNLVVLDETVVTAKNAPVSRKANEVTLGKVV 81
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 62 KVIYEEQIOEQATGRQADVMAQLIPSLG-VSSGTTNFGTMMIG----RQVQFLNGV- 116
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 82 KTA-ETINKEQVNLIRDL-----TRYDPGIAVVEQGRGASSGYIRGMDKIRVAVLVDGIN 136
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 117 -----PLTCSR-DISRLQSLNINQVAVLEVLGSA-TSYVSGCATGLNIVTKS--- 164
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 137 QAQHYQGVVAGKNYAAGGAINIEYENRVSIEISKANSSEYSGALSGSAFVTKTADD 196
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 165 ---DLBEEQFETR-----IGVHSGKLSSEGI-----GYQVG 192
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 197 LKQKQGVQTKTAYASKNNVNSVAAAGKAGSFGSLIIVTDRRGOEYKAHDDAYQGS 256
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 193 QS-----VAGVSENGV-----LARDVDYRTTGGAFDANG-KR 225
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 257 OSFDRAVATTPNNRTFLIANECANGYEACAAAGGQTKLOAKPTNRDKVNVKDYTGPR 316
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 226 IAPQAQTDKQDSKSLSVNTVNDWOLDKQIN--LALTHVNDKQDTPADYDGNRLAVL 283
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 317 LIPNPL--TQDSKSLLRPG--YQNDKHYVGGVYEITKQNYAMQDKTVPAY----LAVH 368
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 284 FGEKPSLN-----AIKGLSLSEQPKTT---KSTFNINHYH 315
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 369 DIEKSLNSHAQANGYQGNLGERIRDTIGDPDSCYGINYAH 410
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 15  
US-08-765-081-5  
; Sequence 5, Application US/08765081  
; Patent No. 5798260  
; GENERAL INFORMATION:  
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.  
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness  
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98101-2347  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage  
; COMPUTER: IBM PC compatible/Pentium  
; OPERATING SYSTEM: MS-Windows 3.1  
; SOFTWARE: Word for Windows-6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/765,081  
; FILING DATE: March 26, 1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/06994  
; FILING DATE: June 7, 1995  
; APPLICATION NUMBER: US 08/265,714  
; FILING DATE: June 24, 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Shelton, Dennis K.  
; REGISTRATION NUMBER: 26,997  
; REFERENCE/DOCKET NUMBER: CHOR-1-10286  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)  
; TELEFAX: 1-206-224-0779  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 696 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

US-08-765-081-5

```

Query Match 6.6%; Score 136; DB 1; Length 696;
Best Local Similarity 21.8%; Pred. No. 0.00028;
Matches 74; Conservative 53; Mismatches 118; Indels 94; Gaps 18;
Qy 40 VITIDKSGMALANRITOMPHTTKVIYEEQIOEQATGRQADVMAQL--IPSLGVSSGTT 97
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 27 VMIYASAG--YEKKLTNAASVSVISQEEIQ-----SSQYHDLAEALRSVEGVDSGTG 79
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 98 SNFG-----QTMHGRQVQFLNGVPLTGSRDIS--RLNSINPN-----QVAREVLGSA 145
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 80 KTGLEISIRGMPASYTLIIIDGVRCQSSDVTNGFSAMTGMPPPLAAIERIEVIRGP 139
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 146 TS-IYSGCATGLNIVTKSDLEEEQFETRIGVH-----GSKLSSEGI 187
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 140 MSTLYGSDMGVGVNIITRKNADKWLSSVAGNLQBSNKGWSSQFNWSSGFLVDDSV 199
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 188 GYQV-----GOSVAGVSENGVNLARLDVDRYTTGAFDANGKRIAPQAQTDKQDSK 239
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 200 SLQVRGSTQQRQSSVTSLS-----DTAGTRI-PYPT-----ESQ 233
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 240 SLSVNTVNDWOLDKQINLALTHYNDKQDTPADYDGNRLAVLFGKPSLSNAIKGLSLS 299
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 234 NYNLGARLDKWKASEQDVL-----WFDMDTIT--RQRYDNR-----DGQLGSLTGGYDRTL- 280
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 300 EQPKTKSTFNINHYHD---DLWGNTINTINAYYRREKGR 335
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 281 ---RYERNKISAGVDHTFTFGTWKSYLNMNE--TENKGR 314
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 16

US-09-098-082-5  
; Sequence 5, Application US/09098082  
; Patent No. 6040421  
; GENERAL INFORMATION:  
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.  
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness  
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98101-2347  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage  
; COMPUTER: IBM PC compatible/Pentium II  
; OPERATING SYSTEM: MS-Windows 95  
; SOFTWARE: Word for Windows-6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/098,082  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/765,081  
; FILING DATE: March 26, 1997  
; APPLICATION NUMBER: PCT/US95/06994  
; FILING DATE: June 7, 1995  
; APPLICATION NUMBER: US 08/265,714  
; FILING DATE: June 24, 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Shelness, Diana K.  
; REGISTRATION NUMBER: 35,356  
; REFERENCE/DOCKET NUMBER: CHOR-1-12402  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)  
; TELEFAX: 1-206-224-0779  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 696 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear

MOLECULE TYPE: protein  
US-09-098-082-5

Query Match 6.6%; Score 136; DB 3; Length 696;  
Best local similarity 21.8%; Pred. No. 0.00028;  
Matches 74; Conservative 53; Mismatches 118; Indels 94; Gaps 18;

```
QY 40 VITIDKSGMALANRITOMPHITKVIYEEOIQEATGSRQLADVAQI--IPSLGVSSGTT 97
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 27 VMTIASAG--YEKLTNMAASVSIVISOELQ-----SSQYHDLAELRSVGVDSGTG 79
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 98 SNFG-----QTMHGRQVQFLNGVPLTGSRDIS--RLQNSINPN-----QVARIIVLSGA 145
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80 KTGGLAISIRGMASVTLILIDVGRQSSSDVTPNGFSAMNTGFMPLAIEIETIRGP 139
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 146 TS-IYSGATGGLINIVTKSPDLEEQFETRIGVH-----GSKLSSEGI 187
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 140 MSTLYSGDAMGAVNIITIRKNADKMLSSVNAAGLQESNKMGNSSQFNFWSSGFLVDDSV 199
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 188 GYQV-----GQSVAGVSENGVNLARLDVYRTTGGAFPDANGRIAPPAQTDKDSK 239
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 200 SLQVRGSGTQGRQSSSVTSLSS-----DTAGTRI-PYPT-----ESQ 233
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 240 SLSTVNTVMDQLDKONINIALTHYNDKQDTDYAPDYGRLAVLFGKPSLNAIKGLSLS 299
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 234 NYNLGARLDWKASEQDVL-----WFDMDTT--RQRYDNR-----DQQLGSLTGGYDRTL- 280
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 300 EOPKTKSTFNINYNHD--DLWGNTINTNAYYRREKGR 335
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 281 ---RYERNKISAGYDHTFTFGTWKSYLNWNE--TENKGR 314
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 17  
PCT-US95-06994-5

Sequence 5, Application PC/TUS9506994

```
GENERAL INFORMATION:
APPLICANT: Children's Hospital & Medical Center
APPLICANT: University of Washington
APPLICANT: Washington State University Research Foundation
APPLICANT: TARR, PHILIP I
APPLICANT: BILGE, SIMA S
APPLICANT: BESSER, THOMAS E
APPLICANT: VARY JR, JAMES C
TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSER: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
STREET: SUITE 2800, 1420 FIFTH AVENUE
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: WA 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: 07-JUN-95
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/265,714
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: BRODERICK, THOMAS F
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: CHOR-18591
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682 8100
TELEFAX: (206) 224 0779
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
```

LENGTH: 696 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-06994-5

Query Match 6.6%; Score 136; DB 5; Length 696;  
Best local similarity 21.8%; Pred. No. 0.00028;  
Matches 74; Conservative 53; Mismatches 118; Indels 94; Gaps 18;

```
QY 40 VITIDKSGMALANRITOMPHITKVIYEEOIQEATGSRQLADVAQI--IPSLGVSSGTT 97
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 27 VMTIASAG--YEKLTNMAASVSIVISOELQ-----SSQYHDLAELRSVGVDSGTG 79
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 98 SNFG-----QTMHGRQVQFLNGVPLTGSRDIS--RLQNSINPN-----QVARIIVLSGA 145
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80 KTGGLAISIRGMASVTLILIDVGRQSSSDVTPNGFSAMNTGFMPLAIEIETIRGP 139
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 146 TS-IYSGATGGLINIVTKSPDLEEQFETRIGVH-----GSKLSSEGI 187
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 140 MSTLYSGDAMGAVNIITIRKNADKMLSSVNAAGLQESNKMGNSSQFNFWSSGFLVDDSV 199
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 188 GYQV-----GQSVAGVSENGVNLARLDVYRTTGGAFPDANGRIAPPAQTDKDSK 239
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 200 SLQVRGSGTQGRQSSSVTSLSS-----DTAGTRI-PYPT-----ESQ 233
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 240 SLSTVNTVMDQLDKONINIALTHYNDKQDTDYAPDYGRLAVLFGKPSLNAIKGLSLS 299
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 234 NYNLGARLDWKASEQDVL-----WFDMDTT--RQRYDNR-----DQQLGSLTGGYDRTL- 280
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 300 EOPKTKSTFNINYNHD--DLWGNTINTNAYYRREKGR 335
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 281 ---RYERNKISAGYDHTFTFGTWKSYLNWNE--TENKGR 314
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

RESULT 18  
PCT-US95-06994-8

Sequence 8, Application PC/TUS9506994

```
GENERAL INFORMATION:
APPLICANT: Children's Hospital & Medical Center
APPLICANT: University of Washington
APPLICANT: Washington State University Research Foundation
APPLICANT: TARR, PHILIP I
APPLICANT: BILGE, SIMA S
APPLICANT: BESSER, THOMAS E
APPLICANT: VARY JR, JAMES C
TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSER: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
STREET: SUITE 2800, 1420 FIFTH AVENUE
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: WA 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: 07-JUN-95
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/265,714
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: BRODERICK, THOMAS F
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: CHOR-18591
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682 8100
```

```
; TELEFAX: (206) 224 0779
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 703 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: E. coli O157:H7 adhesin amino acid sequence,
; DESCRIPTION: wherein "Xaa" residues represent gaps
; DESCRIPTION: introduced to facilitate best alignment with
; DESCRIPTION: SEQ ID NO:9.
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli O157:H7
; STRAIN: 86-24 NALR
PCT-US95-06994-8

Query Match 6.6%; Score 135.5; DB 5; Length 703;
Best Local Similarity 22.4%; Pred. No. 0.00031;
Matches 57; Conservative 43; Mismatches 82; Indels 73; Gaps 12;

Qy 40 VITIDKGMALANRITOMPHTTKVIYEEQIQEQATGSRQLADVMAQL--IPSLGVSSGTT 97
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 32 VMIVSAG--YEKLTNAAASVVISQBELQ-----SQYHDLAELRSVGVDSGTG 84
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 98 SNFG-----QTMHGROVFLNGVPLTGSRDIS-RQLNSINPN-----QVARIIVLSGA 145
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 85 KTGGLGISIRGMPASYTLILIDGVQGGSSDVTNPGFSAMNTGMPPLAAIERIEVIRGP 144
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 146 TS-IVGSGATGLINIVTKSLDEEQPETRIGVH-----GSKLSSEGI 187
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 145 MSTLYGSDAMGVVNIITRKNADKWLSSVNAGLNQESNKGNSQFNFWSSGPLVDDSV 204
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 188 GYQV-----GQSVAGVSENGVNLARLDVYRTTGGAFDANGKRIAPPAQTDKQSK 239
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 205 SLOVRGSGTQORQSSVTSLS-----DTAGTRI-PYPT-----ESQ 238

Qy 240 SLSVNTNVDQLDDK 254
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 239 NVNLGARLDWKASEQ 253
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 19
PCT-US95-06994-6
; Sequence 6, Application PC/TUS9506994
; GENERAL INFORMATION:
; APPLICANT: Children's Hospital & Medical Center
; APPLICANT: University of Washington
; APPLICANT: Washington State University Research Foundation
; APPLICANT: TARR, PHILLIP I
; APPLICANT: BILGE, SIMA S
; APPLICANT: BESSER, THOMAS E
; APPLICANT: VARY JR, JAMES C
; TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
; STREET: SUITE 2800, 1420 FIFTH AVENUE
; CITY: SEATTLE
; STATE: WASHINGTON
; COUNTRY: USA
; ZIP: WA 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: 07-JUN-95
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,714

; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BRODERICK, THOMAS F
; REGISTRATION NUMBER: 31,332
; REFERENCE/DOCKET NUMBER: CHOR-18591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682 8100
; TELEFAX: (206) 224 0779
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 718 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: E. coli O157:H7 adhesin amino acid sequence,
; DESCRIPTION: wherein "Xaa" residues represent gaps
; DESCRIPTION: introduced to facilitate best alignment with
; DESCRIPTION: SEQ ID NO:7.
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli O157:H7
; STRAIN: 86-24 NALR
PCT-US95-06994-6

Query Match 6.6%; Score 134.5; DB 5; Length 718;
Best Local Similarity 24.3%; Pred. No. 0.0004;
Matches 58; Conservative 45; Mismatches 97; Indels 39; Gaps 12;

Qy 40 VITIDKGMALANRITOMPHTTKVIYEEQIQEQATGSRQLADVMAQL--IPSLGVSSGTT 97
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 37 VMIVSAG--YEKLTNAAASVVISQBELQ-----SQYHDLAELRSVGVDSGTG 89
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 98 SNFG-----QTMHGROVFLNGVPLTGSRDIS-RQLNSINPN-----QVARIIVLSGA 145
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 90 KTGGLGISIRGMPASYTLILIDGVQGGSSDVTNPGFSAMNTGMPPLAAIERIEVIRGP 149
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 146 TS-IVGSGATGLINIVTKSLDEEQPETRIGV-----HGSKLSSEGIQVQGVSV 195
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 150 MSTLYGSDAMGVVNIITRKNADKWLSSVNAGLNQESNKGNSQFNFWSSGPLVDDSV 209
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 196 AGVSENGVNLARLDVYRTTGGAFDANGKRIAPPAQTDKQSKLSVNTNVDQLDDK 254
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 210 S-LQVRGSGTQORQSSVXXT-SLSDTAGTRI-PYPT-----ESQYNLGARLDWKASEQ 260
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 20
US-08-613-009A-8
; Sequence 8, Application US/08613009A
; Patent No. 6090576
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Harkness, Robin E
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Du, Run-Pan
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/613,009A
```

```

1      FILING DATE: 08-MAR-1996
2
3      CLASSIFICATION: 435
4
5      ATTORNEY/AGENT INFORMATION:
6
7      NAME: Stewart, Michael I
8      REGISTRATION NUMBER: 24973
9
10     REFERENCE/DOCKET NUMBER: 1038-542
11
12     TELECOMMUNICATION INFORMATION:
13
14     TELEPHONE: (416) 595-1155
15     TELEFAX: (416) 595-1163
16
17     INFORMATION FOR SEQ ID NO: 8:
18
19     SEQUENCE CHARACTERISTICS:
20
21     LENGTH: 1053 amino acids
22     TYPE: amino acid
23     STRANDEDNESS: single
24
25     TOPOLOGY: linear
26
27     JS-08-613-009A-8

```

Query Match	6.5%	Score 134	DB 3	Length 1053
Best Local Similarity	23.4%	Pred. No. 0.00084		
Matches 94	Conservative 48	Mismatches 150	Indels 110	Gaps 21

```

Qy 10 LSLPLSLVA-----VYQOLAQNPESLPPYLEPPVITLIDKSMALANLITOMPHTTKV 63
Db 1 LSLGLLNTIYQALANTTADKABATDKNTLVVVLDEYVYATKAKARK-ANETVJGKXVKT 59
Qy 64 IYEEIOIQOATGSGROLADWMAQOLIPSLG-VSSGTTSNFGOTMGH--ROYOFLINGV--- 116
Db 60 A-ETINKEOVINIRDL-----TRYDPGLAVVEQOGASGSYSIRMDKRWAVAVLDGINQA 114
Qy 117 -----PLTQSR-DISROLINSINPQVARIEVLSGA-TSYSSGATGGLINTVTS--- 164
Db 115 OHVALQGVAGKNTAYAGGAINIEYENRVSYEISKANSSSEYSGALSGSAFVPTKADD 174
Qy 165 ---DLSEQFETR-----IGVHSGKLSSEGI-----GYQVG 192
Db 175 IIKQKMDGVQTKRATVYAKNNNAWNSVAAAGKASFSGLIITYDRGOEYKAHDHAYQGS 234
Qy 193 QS-----VAGVSENGNV-----LARDVDYRTTGAFDPANG-KR 225
Db 235 QSFDRVAVATTPNNRFTFLIANECANGYEAACAAGQTVLQAKPTNVDKRVVKDYTPNR 294
Qy 226 IAPBAQTDKQDSKLSVNTVDWQLODKONIN--LALTHTNBDKQDTDYAPDYANRLAVL 263
Db 295 LIPNPL---TQDSKSLLRPG--YQJLNDKHYVGVGEYETITKONYAMODKVPAY---LTVH 346
Qy 284 FGEXPSLN-----AIKGLSLSEOPKTT---KSTFNINVYH 315
Db 347 DIEKSRITNMAOANGYQYGNNLGEBIRITPTIPDGGYGINVAH 388

```

RESULT 21  
 US-08-778-5708-10  
 Sequence 10 Application US/0878570B  
 Patent No. 6437096  
 GENERAL INFORMATION:  
 APPLICANT: Myers, Lisa E  
 APPLICANT: Schryvers, Anthony B  
 APPLICANT: Harkness, Robin E  
 APPLICANT: Loosmore, Sheena M.  
 APPLICANT: Du, Run-Pan  
 APPLICANT: Yang, Yan-Ping  
 APPLICANT: Klein, Michel H  
 TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
 NUMBER OF SEQUENCES: 43  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Slim & McBurney  
 STREET: 6th Floor, 330 University Avenue  
 CITY: Toronto  
 STATE: Ontario  
 COUNTRY: Canada  
 ZIP: M5G 1R7  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.10
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/778,570B
FILING DATE: 03-JAN-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24972
REFERENCE/DOCKET NUMBER: 1038-664
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1053 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IS-08-778-570B-10

```

Query Match	6.5%	Score	134	DB	4	length	1053
Best Local Similarity	23.4%	Pred. No.	0.00084				
Matches	94	Conservative	48	Mismatches	150	Indels	110
						Gaps	21

```

0Y 10 LSTLLELSVA-----VVOQLADQNESELPVELEPVVILIDKSGMALANKKIQMPHTIKV 63
Db 1 LSLGLINTQVALANTTADKAEATDKTNLVVLDELFTVATKQARK-ANENVTGKGVKVT 59
QY 64 IYEOIOEQAGSGROLADWMAQLPSLG-VSSGTSNFGQTHG---ROYQPLNCGV--- 116
Db 60 A-ETINKEOVNLINDL-----TRYDPGIAVVEOGKASSGYSIGMDKNRAVAVLVDGINQA 114
QY 117 -----PLTCSR-DISROLNSINPNQVAREVLSGA-TSYSGATGGLINITYKS--- 164
Db 115 QHVALQGPVAGKNTAAAGALNIEIENVRSEVLSKGNSSSEYSGALLSGSVAFVTKTADD 174
QY 165 ---DLEBQFETR-----IGVHSGKLSSEGI-----GYOVC 192
Db 175 IIKKQKMGVQTKTAAVAYSKNNAMVNSVAAAGKGSFSGILIIYTRDRCQEYKAHDADVOGS 234
QY 193 QS-----VAGVSENGAV-----LARLDVDFYTTGCAFPANG-KR 225
Db 235 QSPRAVATTPNNRTFELIANECANGVEYCAAGAGQTKLQAKPTNVAVDKNVVNDYTGPNR 294
QY 226 IAPBPAQCDKRODSLSLVNTNVMDQLDDKONIN-LALTNYNDKQDTPVAPDYGKRLAVL 283
Db 295 LIPNPL--TQDSKSLLRPLRG-YQLNDKHYGVGVYEITKONYAMODKTVPAY---LTVH 346
QY 284 FGEKPSLN-----AIKGLSLEQPKTT---KSTFNINHH 315
Db 347 DIBKSRUSNHAQANGYYQANNLGERIRIDTTPGDSGYGYNIAH 388

```

RESULT 22 584-10  
 US-09-059-584-10  
 : Sequence 10, Application US/09059584  
 : Patent No. 6440701  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: Myers, Lisa E  
 : APPLICANT: Schryvers, Anthony B  
 : APPLICANT: Harkness, Robin E  
 : APPLICANT: Loosmore, Sheena M.  
 : APPLICANT: Du, Run-Pan  
 : APPLICANT: Yang, Yan-Ping  
 : APPLICANT: Klein, Michel H  
 : TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
 : NUMBER OF SEQUENCES: 60  
 :  
 : CORRESPONDENCE ADDRESSES:  
 : ADDRESSEE: Sim & McBurney  
 : STREET: 6th Floor, 330 University Avenue  
 : CITY: Toronto

```

; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/059,584
; FILING DATE: 14-APR-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/778,570
; FILING DATE: 03-JAN-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-794
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1053 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-059-584-10

```

```

Query Match      6.5%; Score 134; DB 4; Length 1053;
Best Local Similarity 23.4%; Pred. No. 0.0084;
Matches 94; Conservative 48; Mismatches 150; Indels 110; Gaps 21;

```

```

QY 10 LSLPLLSVA-----VTQOLYAQPNEISLPTVELEPVPVITIDKSGMALANRITQMPHTTKV 63
DB 1 LSLGLLNTQVALANTADKAEATDKTNLVVLDVETVTTAKKNARK-ANEVTGLGKVVKT 59
QY 64 IYEQIQEATGSQLADVMAGLIPSLG-VSSGTTTSNFGQTMHG---RQVQFLNGV--- 116
DB 60 A-ETINKEQVLNIRD-----TRYDPCIIVVEQGRGASSGYSIRGMDKNRVAVLVDGINQA 114
QY 117 -----PLTGRS-DISRLNSINPNQVARIIVLSGA-TSIYSGATGGLINIVTKS--- 164
DB 115 QHYALQGPVAGKNYAAGGAINIEYENVRSEISKGANSSEYSGSALSGSAFVTKTADD 174
QY 165 ---DLEEEQFETR-----IGVHGSKLSSEGI-----GYQVG 192
DB 175 IIKGKDMGVQTKTAYASKNNAWNSVAAAGKAGSFGSLIITDRRGQYKAHDDAYOGS 234
QY 193 QS-----VAGVSENGNV-----LARLDVDYRTTGGAFDANG-KR 225
DB 235 QSFDRAVATDPNNRTFLIANECANGNYEACAAGGQTKLQAKPTNVRDKVNVKDYTGPNR 294
QY 226 IAPPAQTDKQSKLSVNTVNDWQLDDKQIN--LALTHYNDKQDTPDYAPDYGRLAVL 283
DB 295 LIPNPL---TQDSKSLLRPG--YQLNDKHYVGGVYEITKQNYAMQDKTVPAY---LTVH 346
QY 284 FQEKPSLN-----AIKGLSLSEQPKTT---KSTFNINYHH 315
DB 347 DIEKSRSLNHAQANGYQGNLGERIRDTIGPDSGYGINYAH 388

```

```

RESULT 23
US-08-613-009A-7
; Sequence 7, Application US/08613009A
; Patent No. 6090576
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryver, Anthony B
; APPLICANT: Harkness, Robin E
; APPLICANT: Loosmore, Sheena M.

```

```

; APPLICANT: Du, Run-Pan
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/613,009A
; FILING DATE: 08-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-542
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1074 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-613-009A-7

```

```

Query Match      6.5%; Score 134; DB 3; Length 1074;
Best Local Similarity 23.4%; Pred. No. 0.0087;
Matches 94; Conservative 48; Mismatches 150; Indels 110; Gaps 21;
QY 10 LSLPLLSVA-----VTQOLYAQPNEISLPTVELEPVPVITIDKSGMALANRITQMPHTTKV 63
DB 22 LSLGLLNTQVALANTADKAEATDKTNLVVLDVETVTTAKKNARK-ANEVTGLGKVVKT 80
QY 64 IYEQIQEATGSQLADVMAGLIPSLG-VSSGTTTSNFGQTMHG---RQVQFLNGV--- 116
DB 81 A-ETINKEQVLNIRD-----TRYDPCIIVVEQGRGASSGYSIRGMDKNRVAVLVDGINQA 135
QY 117 -----PLTGRS-DISRLNSINPNQVARIIVLSGA-TSIYSGATGGLINIVTKS--- 164
DB 136 QHYALQGPVAGKNYAAGGAINIEYENVRSEISKGANSSEYSGSALSGSAFVTKTADD 195
QY 165 ---DLEEEQFETR-----IGVHGSKLSSEGI-----GYQVG 192
DB 196 IIKGKDMGVQTKTAYASKNNAWNSVAAAGKAGSFGSLIITDRRGQYKAHDDAYOGS 255
QY 193 QS-----VAGVSENGNV-----LARLDVDYRTTGGAFDANG-KR 225
DB 256 QSFDRAVATDPNNRTFLIANECANGNYEACAAGGQTKLQAKPTNVRDKVNVKDYTGPNR 315
QY 226 IAPPAQTDKQSKLSVNTVNDWQLDDKQIN--LALTHYNDKQDTPDYAPDYGRLAVL 283
DB 316 LIPNPL---TQDSKSLLRPG--YQLNDKHYVGGVYEITKQNYAMQDKTVPAY---LTVH 367
QY 284 FQEKPSLN-----AIKGLSLSEQPKTT---KSTFNINYHH 315
DB 368 DIEKSRSLNHAQANGYQGNLGERIRDTIGPDSGYGINYAH 409

```

```

RESULT 24
US-08-778-570B-9
; Sequence 9, Application US/08778570B
; Patent No. 6437096

```



```
GENERAL INFORMATION:
APPLICANT: Myers, Lisa E
APPLICANT: Schryvers, Anthony B
APPLICANT: Hartness, Robin E
APPLICANT: Loomore, Sheena M.
APPLICANT: Du, Run-Pan
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/778,570B
FILING DATE: 03-JAN-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24973
REFERENCE/DOCKET NUMBER: 1038-664
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1074 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-778-570B-9

Query Match
Best Local Similarity 23.4%; Score 134; DB 4; Length 1074;
Matches 94; Conservative 48; Mismatches 150; Indels 110; Gaps 21;

10 LSLPLLSVA-----VTQQLVAQPNESLPTVELPVIITIDKSGMALANRITOMPHTTV 63
22 LSLGLNITVOVALANTTADKAETDKTNLVVLDDEVTTAKKNARK-ANEVTLGKVVKT 80
64 IYEEQIOEQATGSRQADVAQQLPSLG-VSSGTTSNFGQTMHG---RQVQPLNGV--- 116
81 A-ETINKEQVLNIRD-----TRYDPIGAVVEQGRGSSGYSIRGMDKNRVAVLVDGINQA 135
117 -----PLTGR-DISROLNSINPQVARIYVLSGA-TSIYSGATGGLINIVTKS--- 164
136 OHYALOGPVAQKNYAAGAINIEYENRVSVEISKANSSEYSGALSGSAFVTKTAD 195
165 -----DLEEBQFETR-----IGVHGSGLSSEGI-----GYQVG 192
196 IIKDGKDWGVQTKTAYASKNNAVNSVAAGKAGSPSGLIITDRRGQRYKAHDADAYQGS 255
193 QS-----VAGVSENGNV-----LARDLVYRTTGAFPDANG-KR 225
256 QSFDAVAVATTDPNNRTFLIANECANGNYEACAAGGQTKQAQKPTNVRDKNVAVDTGPNR 315
226 IAPBPQTKDKQSKLSVNTNVDWQJDDKONIN--LALTNYNDKQDTYAPDYGNRLAVI 283
316 LIPNPL-----TQDSKSLILBPG--YQLNDKHVGVGVETIKQNYAMQDKTVPAY---LTVH 367
284 FGKPSLN-----AIKGLSLSEQPKT---KSTFNINHH 315
368 DIKSRLSNHAQANGYYQGNLGERIRDTTIGPDGGINYAH 409
```

```
RESULT 25
US-09-059-584-9
Sequence 9, Application US/09059584
Patent No. 6440701
GENERAL INFORMATION:
APPLICANT: Myers, Lisa E
APPLICANT: Schryvers, Anthony B
APPLICANT: Hartness, Robin E
APPLICANT: Loomore, Sheena M.
APPLICANT: Du, Run-Pan
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/059,584
FILING DATE: 14-APR-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/778,570
FILING DATE: 03-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24973
REFERENCE/DOCKET NUMBER: 1038-794
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1074 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-059-584-9

Query Match
Best Local Similarity 23.4%; Score 134; DB 4; Length 1074;
Matches 94; Conservative 48; Mismatches 150; Indels 110; Gaps 21;

10 LSLPLLSVA-----VTQQLVAQPNESLPTVELPVIITIDKSGMALANRITOMPHTTV 63
22 LSLGLNITVOVALANTTADKAETDKTNLVVLDDEVTTAKKNARK-ANEVTLGKVVKT 80
64 IYEEQIOEQATGSRQADVAQQLPSLG-VSSGTTSNFGQTMHG---RQVQPLNGV--- 116
81 A-ETINKEQVLNIRD-----TRYDPIGAVVEQGRGSSGYSIRGMDKNRVAVLVDGINQA 135
117 -----PLTGR-DISROLNSINPQVARIYVLSGA-TSIYSGATGGLINIVTKS--- 164
136 OHYALOGPVAQKNYAAGAINIEYENRVSVEISKANSSEYSGALSGSAFVTKTAD 195
165 -----DLEEBQFETR-----IGVHGSGLSSEGI-----GYQVG 192
196 IIKDGKDWGVQTKTAYASKNNAVNSVAAGKAGSPSGLIITDRRGQRYKAHDADAYQGS 255
193 QS-----VAGVSENGNV-----LARDLVYRTTGAFPDANG-KR 225
256 QSFDAVAVATTDPNNRTFLIANECANGNYEACAAGGQTKQAQKPTNVRDKNVAVDTGPNR 315
```

Qy	226	IAREPAQTDKQDSKSLSVNTVNDWOLDKQIN--LALTHYNDKQDQTDVAPDYGNELAVL	283
Db	316	LLENPL---TQDSKSLLRFG--YQLNDRHYVGGVVEITKQNYAMQDKTVPAY---LTVH	367
Qy	284	FGKPSLN-----AIKGLSLSEQPKTT---KSTFNINVHH	315
Db	368	DIKSRSLNHAOANGYYOGNNGERDRTTIGPDSGYINVAH	409

RESULT 26

```

US-08-425-843--7
; Sequence 7, Application US/08425843
; Patent No. 6020154
; GENERAL INFORMATION:
; APPLICANT: Hansen, Eric J.
; APPLICANT: Cope, Leslie D.
; APPLICANT: Jacobik, Gregory P.
; APPLICANT: Hanson, Mark S.
; TITLE OF INVENTION: H. influenzae HxB and HxC Genes, Proteins
; TITLE OF INVENTION: and Methods of Use
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210-4433
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/425,843
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: AMCY:012/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 715 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-425-843-7

```

Db	160	GV	--KIROQ--GYQQTANLIS--ERDASVPAAND-KFDVLISAFYNNADNL-----RTGK	205
QY	236	QDS	-----KSLSVNTNVMDQLDDKQINIALTHYNDKQDQTDYAPDYGNRVLAVFGEKPS	2899
Db	206	GNKLNN	TA YKQFGLAKFGWQINANDNRVLSHRETFKQTA-----PS	248
QY	290	LNAIKGLS	LESEQPKTTKSTFNINHYHDDLWG-----NTINNAVYRREKGRF	336
Db	249	NNEVNE	ELTNE-----KIDQNEFGNSNGNLAPORAKPSSETSAFYSKVKTRF	296

## RESULT 27

US-08-778-570B-14  
 ; Sequence 14, Application US/08778570B  
 ; Patent No. 6437096  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Myers, Lisa E  
 ; APPLICANT: Schryvers, Anthony B  
 ; APPLICANT: Harkness, Robin E  
 ; APPLICANT: Loosmore, Sheena M.  
 ; APPLICANT: Du, Run-Pan  
 ; APPLICANT: Yang, Yan-Ping  
 ; APPLICANT: Klein, Michel H  
 ; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
 ; NUMBER OF SEQUENCES: 43  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Sim & McBurney  
 ; STREET: 6th Floor, 330 University Avenue  
 ; CITY: Toronto  
 ; STATE: Ontario  
 ; COUNTRY: Canada  
 ; ZIP: M5G 1R7  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/778,570B  
 ; FILING DATE: 03-JAN-1997  
 ; CLASSIFICATION: 536  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Stewart, Michael I  
 ; REGISTRATION NUMBER: 24973  
 ; REFERENCE/DOCKET NUMBER: 1038-664  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (416) 595-1155  
 ; TELEFAX: (416) 595-1163  
 ; INFORMATION FOR SEQ ID NO: 14:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1052 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-08-778-570B-14

Db 175 IKDKDGVQOTKTAYASKNNAWNSVAAAGKAGSFGSLITTYDRRGQERYKAHDADYQGS 234  
Qy 193 QS-----VAGVSENGNV-----LARDVDYRTTGAFDANG-KR 225  
Db 235 QSFDRAVATTDPPNPKFLIANECANGYEACAGGQTKLOAKPTNVBDKVNVDYGPNR 234  
Qy 226 IAPPAQTDKQDSKSLSVNTNVDMQDDKONIN--LALTHYNDKQDTPVAPDYGNRLAVL 283  
Db 295 LIENPL---TQDSKSLILRPG--YQINDKHVGGVYEITQONVAMQDKTVPAV---LTVH 346  
Qy 284 FGKPSLN-----AIKGLSLSEQPKT---TKSTFNINVHH 315  
Db 347 DIEKSLSHNGANGYQGNLGERIRDAIGANSYGGINYAH 388

RESULT 28  
US-09-059-584-14  
; Sequence 14, Application US/09059584  
; Patent No. 6440701

; GENERAL INFORMATION:  
; APPLICANT: Myers, Lisa E  
; APPLICANT: Schryvers, Anthony B  
; APPLICANT: Harkness, Robin E  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Du, Run-Pan  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
; NUMBER OF SEQUENCES: 60  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/059,584  
; FILING DATE: 14-Apr-1998  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/778,570  
; FILING DATE: 03-JAN-1997  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-794  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1052 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-09-059-584-14

Query Match 6.4%; Score 130; DB 4; Length 1052;  
Best Local Similarity 23.1%; Pred. No. 0.002;  
Matches 93; Conservative 48; Mismatches 151; Indels 110; Gaps 21;

Qy 10 LSLPLSLVA-----VQQLYAQPNESLPVLEPVITIDKSGMALANRTQMPHTTKV 63  
Db 1 LSLGLNITGVALLANTADRAEATDKTNLVVLDVETVTAKGNARK-ANEVTLGKRVKVT 59

Qy 64 IYEQIQEQATSGRQIADVMAQLIPSLG-VSSGCTSNFGQTMHG---ROYQFLNGV--- 116  
Db 60 A-ETINKEQVLNIRDL---TRYDPIAVVEQGRGSSGYSIRGDKXNVAVLVGDIQA 114  
Qy 117 -----PLTGR-DISRLNSINPQVARIIVLGA-ITISGATGGLINIVKRS--- 164  
Db 115 QHYALOGPAVGNKYAAGAGINEIEYENVSVEISKGANSSEYSGALSGSAFVTKTAD 174  
Qy 165 ---DLEEQFETR-----IGVHGSKLSEGI-----GYGVG 192  
Db 175 IKDKDGVQOTKTAYASKNNAWNSVAAAGKAGSFGSLITTYDRRGQERYKAHDADYQGS 234  
Qy 193 QS-----VAGVSENGNV-----LARDVDYRTTGAFDANG-KR 225  
Db 235 QSFDRAVATTDPPNPKFLIANECANGYEACAGGQTKLOAKPTNVBDKVNVDYGPNR 234  
Qy 226 IAPPAQTDKQDSKSLSVNTNVDMQDDKONIN--LALTHYNDKQDTPVAPDYGNRLAVL 283  
Db 295 LIENPL---TQDSKSLILRPG--YQINDKHVGGVYEITQONVAMQDKTVPAV---LTVH 346  
Qy 284 FGKPSLN-----AIKGLSLSEQPKT---TKSTFNINVHH 315  
Db 347 DIEKSLSHNGANGYQGNLGERIRDAIGANSYGGINYAH 388

RESULT 29  
US-08-613-009A-11  
; Sequence 11, Application US/08613009A  
; Patent No. 6090576

; GENERAL INFORMATION:  
; APPLICANT: Myers, Lisa E  
; APPLICANT: Schryvers, Anthony B  
; APPLICANT: Harkness, Robin E  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Du, Run-Pan  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/613,009A  
; FILING DATE: 08-MAR-1996  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-542  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1070 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-613-009A-11

Query Match 6.4%; Score 130; DB 3; Length 1070;  
Best Local Similarity 23.1%; Pred. No. 0.0021;  
Matches 93; Conservative 48; Mismatches 151; Indels 110; Gaps 21;

```

Qy 10 LSLPLLSVA-----VTQOLYAQNESLPTVELEPVVITIDKSGMALANRITQMPHTTKV 63
Db 19 LSLGLLNIQVALANTTADKAEATDKTNLVVLDVTVTAKKNARK-ANEVTGLGKVVKT 77
Qy 64 IYEEQIOEATGSRQADVMAQLIPSLG-VSSGTTSNFGOTMHG---ROVQFLNGV--- 116
Db 78 A-ETINKEOVLNIRD-----TRYDPGIAVVEQGRGASSYSIRGMDKNRVAVLVDGINQA 132
Qy 117 -----PLTGSR-DISRLNSINPNQVARIIVLSGA-TSIYSGATGGLINIVTKS--- 164
Db 133 QHYALQGVAGKNYAAGGAINIEVENRVSVEISKANSSEYSGSALSGSVAFVTKTADD 192
Qy 165 ---DLEBEOFETR-----IGVHGSKLSSGI-----GYQVG 192
Db 193 IIKDGMGVQTKTAYASKNNWVNSVAAAGKAGSFGSLIITDRRGQEKYKAHDDAYQGS 252
Qy 193 QS-----VAGVSENGNV-----LARLDVDYRTTGGAFDANG-KR 225
Db 253 QSFDRAVATDPNNPKFLIANECANGNYEACAGGOTKLOAKPTNVRDKVNVKDYTGPNR 312
Qy 226 IAPPAOTDKQDSKLSVNTNVDQDKQININ--LALTHYNDKQDTPDYAPDYGNNRLAVL 283
Db 313 LIPNPL---TQDSKSLLRPG--YQNDKHVYGVYEITKQNYAMQDKTVPAY---LTVH 364
Qy 284 FGEKPSLN-----AIKGLSLEQPKT---TKSTFNINHH 315
Db 365 DIEKRLSNHGQANGYQGNLGERIRDAIGANSYGINYAH 406

```

RESULT 30

```

US-08-778-570B-13
; Sequence 13, Application US/08778570B
; Patent No. 6437096
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Harkness, Robin E
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Du, Run-Pan
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 03-JAN-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-664
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1070 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

US-08-778-570B-13

```

Query Match 6.4%; Score 130; DB 4; Length 1070;
Best Local Similarity 23.1%; Pred. No. 0.0021;
Matches 93; Conservative 48; Mismatches 151; Indels 110; Gaps 21;
Qy 10 LSLPLLSVA-----VTQOLYAQNESLPTVELEPVVITIDKSGMALANRITQMPHTTKV 63
Db 19 LSLGLLNIQVALANTTADKAEATDKTNLVVLDVTVTAKKNARK-ANEVTGLGKVVKT 77
Qy 64 IYEEQIOEATGSRQADVMAQLIPSLG-VSSGTTSNFGOTMHG---ROVQFLNGV--- 116
Db 78 A-ETINKEOVLNIRD-----TRYDPGIAVVEQGRGASSYSIRGMDKNRVAVLVDGINQA 132
Qy 117 -----PLTGSR-DISRLNSINPNQVARIIVLSGA-TSIYSGATGGLINIVTKS--- 164
Db 133 QHYALQGVAGKNYAAGGAINIEVENRVSVEISKANSSEYSGSALSGSVAFVTKTADD 192
Qy 165 ---DLEBEOFETR-----IGVHGSKLSSGI-----GYQVG 192
Db 193 IIKDGMGVQTKTAYASKNNWVNSVAAAGKAGSFGSLIITDRRGQEKYKAHDDAYQGS 252
Qy 193 QS-----VAGVSENGNV-----LARLDVDYRTTGGAFDANG-KR 225
Db 253 QSFDRAVATDPNNPKFLIANECANGNYEACAGGOTKLOAKPTNVRDKVNVKDYTGPNR 312
Qy 226 IAPPAOTDKQDSKLSVNTNVDQDKQININ--LALTHYNDKQDTPDYAPDYGNNRLAVL 283
Db 313 LIPNPL---TQDSKSLLRPG--YQNDKHVYGVYEITKQNYAMQDKTVPAY---LTVH 364
Qy 284 FGEKPSLN-----AIKGLSLEQPKT---TKSTFNINHH 315
Db 365 DIEKRLSNHGQANGYQGNLGERIRDAIGANSYGINYAH 406

```

RESULT 31

```

US-09-059-584-13
; Sequence 13, Application US/09059584
; Patent No. 6440701
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Harkness, Robin E
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Du, Run-Pan
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 14-APR-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/778,570
; FILING DATE: 03-JAN-1997
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-794

```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1070 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-059-584-13

```

```

Query Match      6.4%; Score 130; DB 4; Length 1070;
Best Local Similarity 23.1%; Pred. No. 0.0021;
Matches 93; Conservative 48; Mismatches 151; Indels 110; Gaps 21;

```

```

QY 10 LSLPLSLVA-----VYQOLYAGPNSLPTVELEPVITIDKSGMALANRITQMPHTTKV 63
DB 19 LSLGLNITVOALNANTADRAEATDKNLVVLDLTVTAKGNARK-ANEVTLGKRVKT 77
QY 64 IYERQIQEQATGSRQLADVMAQLIPSLG-VSSGTTNFGQTMHG---ROYQFLINGV--- 116
DB 78 A-ETINKEQVLANIRDL-----TRVPGIAYVVGQGRGASSGISIRGMDKNRAVAVDGINQA 132
QY 117 -----PLTGR-DSRQLNSINPNQVARIELVLSGA-TSISGATGGLINIVTKS--- 164
DB 133 QHVALQGPVAGKNYAAGALNIEIYENVRSVEISKANSSEYSGALGSVAFTXTADD 192
QY 165 ---DLEEQEETR-----IGVHSGKLSSEGI-----GYQVG 192
DB 193 IIKDKQWGVQTKTAYASKNNAWNSVAAGKAGSFSGLLIYTRRGOEYKADDAQGS 252
QY 193 QS-----VAGVSENGNV-----LARDVYRTTGAFDPANG-KR 225
DB 253 QSPRAVAITDPNNPKFLIANECANGYEAACAAGQTKLQAKPNNVADKVVAVDYGPNR 312
QY 226 IAPPAQTDKQDSKLSVNTNVMDQLDKQIN--LALTHYNDKQDTPADPDYGNRLAVL 283
DB 313 LIPMPL---TQDSKSLLRPG--YQLNDKHVGVGVETTKQNVAMQDKTVAVY---LTVH 364
QY 284 FGKPSLN-----ATKGLSLSRQPKT---TKSTPNINTHH 315
DB 365 DIKSRLSNHQANGYQGNLGRIRDAIGANGSYGINVAH 406

```

## RESULT 32

```

US-08-537-361E-8
; Sequence 8, Application US/08537361E
; GENERAL INFORMATION:
; PATENT No. 6121037
; APPLICANT: Stojiljkovic, Igor
; APPLICANT: So, Magdalene
; APPLICANT: Hwa, Vivian
; APPLICANT: Heffron, Fred
; APPLICANT: Nasbif, Xavier
; TITLE OF INVENTION: No. 6121037el Bacterial Hemoglobin Receptor
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/537.361E
; FILING DATE: 02-Oct-1995
; CLASSIFICATION: 536

```

```

; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6121037nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 94,784-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 790 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULAR TYPE: protein
US-08-537-361E-8

```

```

Query Match      6.3%; Score 128; DB 3; Length 790;
Best Local Similarity 23.4%; Pred. No. 0.002;
Matches 95; Conservative 44; Mismatches 151; Indels 116; Gaps 22;

```

```

QY 12 LPLSLVAVTQQLYAQP-----NESLPT-VELEPVITIDKSGMALANRITQMPHTTK 62
DB 7 LPL--AALVGSIFGNPVLADEAATETTPVKAIEKVRVMDQLNAPATVERV-----NLG 59
QY 63 VIYEQIQEQATGSRQLADVMAQLIPSLGVSGTTSNFG---QTMGRQVQFLINGVPLT 119
DB 60 RIQCEMIRDNKDLVRYSTDV-----GLSDSGRHQKGFVAGVGEGRVGSIDGSLP 111
QY 120 GSRDIS---RQLN-----SINPNQVARIELVLSGATSI-YSGATGGLINIVTKSD---- 165
DB 112 DSRNSLYAAYGNPNSSRLSIDPELVNRIEIAKGAOSFNTGSALGGVAVYQTLQGHDL 171
QY 166 LEEBOF-----ETR-----IGVHSGKLSSEGI-GYQVGSVAGVSENGVNLAR 207
DB 172 LDDRQFQVMKNGYSTRNREWTNLGFGVSNDRVDALLYSQRGHETESAGRG----- 226
QY 208 LDVYRTTG---GAPDANGRIAPPAQTDKQDSKLSVNTNVMDQLDKQINLALTHY 264
DB 227 ---YPEVAGSAGALLIRGSSRGIPDPSEK-----HKYNFLKXIAQIINDGHRIG---PSF 274
QY 265 NDKQDDY-----APDYGNRL-AVLPGE-KPSLNALKGLSLSRQPKTK- 306
DB 275 NGQGHNYTIEESYNTLASSWREADDVNRNRNANLFEWTPDSNWSLSLADPDYQTKV 334
QY 307 -----STFNINYHDDLWGNTINTNAAYYRREKGRFYFP 339
DB 335 AAYVKNKSPPTDYSTLIRYNQKDL-----ENIYRSMQTRFRFR 374

```

## RESULT 33

```

US-08-817-707-8
; Sequence 8, Application US/08817707
; PATENT No. 6277382
; GENERAL INFORMATION:
; APPLICANT: Stojiljkovic, Igor
; APPLICANT: So, Magdalene
; APPLICANT: Hwa, Vivian
; APPLICANT: Heffron, Fred
; APPLICANT: Nasbif, Xavier
; TITLE OF INVENTION: No. 6277382el Bacterial Hemoglobin Receptor
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817.707
; FILING DATE: 19-AUG-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6277382nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 94,784-J
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 790 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-817-707-8

Query Match      6.3%; Score 128; DB 4; Length 790;
Best Local Similarity 23.4%; Pred. No. 0.002;
Matches 95; Conservative 44; Mismatches 151; Indels 116; Gaps 21;

Qy 12 LPLLSVAVTQQLYAQP-----NESLPT-VELEPVVITIDKSGMALANRITQPHPTTK 62
Db 7 LPI--AALVGSIFGNPVLAADEAATETTPVKABIKVRYKQQLNAPATVERV-----NLG 59

Qy 63 VIYEQIOEQATGSRQLADVMAQLIPSGVSGTTSNFG---QTMHGRQVQPLNGVPLT 119
Db 60 RIQEMIRNDKDLVRYSTDV-----GLSDSRHQKGFVARGVGEGRVGVSDIGVSLP 111

Qy 120 GSRDIS---RLN-----SINPNQVAREIVLSGATSI-YGSGATGGLINIVTKSD----- 165
Db 112 DSEENSLVARYGNFNSSRLSIDPELVRLNIEIAKGADSENTGSGALGGVYVQTLQCHDLL 171

Qy 166 LEEQFETFRIGVHGSKLSE---GLGYQV-----CQSVAGVSENGVNLAR 207
Db 172 LDDROFGVMKNGYSRRNREWTNTLGFVSDNRVDAALLYSQRRHETESAGERG----- 226

Qy 208 LDVDYRTTG---GAPDANGKRIAPPAQTDKQDSLSLVNTVDMQLDDKQINILALTHY 264
Db 227 ---YFVGGAGSGLIRGSSRGIPDSK-----HKYHFLGKIAYQINDKXHRIG---PSF 274

Qy 265 NDKQDPTY-----APDYGNRL-AVLFGK-KPSLNAIKGLSLSEOPKTKK- 306
Db 275 NGOQGHNTIEESYNLTASSWREADDVNRRRNANLFYEWTPDSNWLSSLKADFDYQTTKV 334

Qy 307 -----STFNINYHDDLWGNTINTNAYYRREKGRFYFP 339
Db 335 AAVNNKGSFPTDYSTWTRYNQKDL-----ENIYNRSMDTRFKRP 374

RESULT 34
US-08-867-941-23
; Sequence 23, Application US/08867941
; Patent No. 5977337
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Du, Run-Pan
; APPLICANT: Wang, Quijun
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: LACTOPERRIN RECEPTOR GENES OF MORAXELLA
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/867,941
; FILING DATE: 03-JUN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-681 MIS:jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 944 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-867-941-23

Query Match      6.2%; Score 127; DB 2; Length 944;
Best Local Similarity 21.1%; Pred. No. 0.0033;
Matches 103; Conservative 50; Mismatches 150; Indels 186; Gaps 22;

Qy 1 MRHSYFQWLSPLLSVAVTQQLY-AQPNESLP-----TVELEPVVITIDKSGMALANRIT 55
Db 1 MNKXHQGQ-LTTLALAVAAAPPVYAANPETAAPDAQTSQSLKEVTYRAAKVGRSKEAVT 59

Qy 56 QMPHTTKVIVVEEQIOEQATGSRQLADVMAQLIPSLG-VSGTTSNFGQTMHG---RQVQF 111
Db 60 GLGKIATK-SETLNKEQVLGIRDL---TRYDPGVAVVVEQNGASGGYSIRGVDKNRVAV 114

Qy 112 LLNGVPLT-----GSRDISRLNSINPNQVAREIVLSGA-TSIYSGATGGLI 158
Db 115 SVDGVAQIQAFVQGSLSGYGGRGGGAINIEIYENISTVEIDKGGAGSDHGGALGGAV 174

Qy 159 NIVTK---SDL-----EEQFETRIG----- 176
Db 175 AFRTKEAADLISDGKSWGIOAKTAYGSKNQPFMSLGSANGFSGKDGWEGLLITERQGRETH 234

Qy 177 -----VHGSKLSSEGIQYQV---GOS-----VAGVSENGVNL- 205
Db 235 PHGDIADGVAYGINRLDAFRQTYGIKKPSEGGYFLAEGSELKPVAKVAGNGVYLNQL 294

Qy 206 -----ARLDVDYRTTGAFDANGKRIAPPAQTDKQ 236
Db 295 NRWVKERIEQNPQLSABEEAMVREAQARHENLSAQAYTGGG-----RILPDP-----M 342

Qy 237 DSKSLSVNTNVDMQL-----DDKQINILALTHYNDKQDTPDYAPDYGNRLAVLFG 285
Db 343 DYRSGSLAKLGYRFGGRHYGVGFEDTKQRYDIRDMTEKQYGTDEAKKFRDSGVYDG 402

Qy 286 E-----XPSLNAIKGLSLSEOPKTKKSTFNINYHDDLWGNTINTNAYYR 330
Db 403 DDFRDGLYFVFNIEEWKQDKLIRGIGL-----KYSRTKF-IDEHH-----RR 444

Qy 331 REKGRFYFP 339
Db 445 RMGLLYRY 453

RESULT 35
US-09-074-658-23
; Sequence 23, Application US/09074658
; Patent No. 6184371
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Run-Pan Du
; APPLICANT: Quijun Wang
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H

```

```

; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA
; NUMBER OF SEQUENCES: 78
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,658
; FILING DATE: 08-MAY-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 944 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-074-658-23

```

```

Query Match      6.2%; Score 127; DB 4; Length 944;
Best Local Similarity 21.1%; Pred. No. 0.0033;
Matches 103; Conservative 50; Mismatches 150; Indels 186; Gaps 22;

```

```

QY 1 MRSHYFQMLSLPLLSAVVQOLY-AQPNESLP-----TVELEPVVITIDSGMALARIT 55
DB 1 MNKHGQ-LITLTAIVAAAPSYAANPETAAPDAOTOSLKEVTAAAKVGRSKAVT 59
QY 56 QMPHTTVIIEEOIQEATSGROLADYMAQLPSLG-VSSGTTSNFGQTMHG--ROVOF 111
DB 60 GLGKIART-SETLNKEQVLGIRDL---TRYDPGVAVVEQNGASGVYIRGVQKNVAV 114
QY 112 LINGVPLT-----GSRDISROLNSINNOVARIEVLSCA-TSIYSGATGGLI 158
DB 115 SVGVAAQIOAFTVQSGISGYGGRGSGAINIEIYENISTVEIDHAGSSDHSGALGAV 174
QY 159 NIVTK--SDI-----EBEOPETRIG----- 176
DB 175 APTKKAADLISDGKMGIOAKTAYGSKNROFMKSLDAGSKDGEGLLRTTEQGHETH 234
QY 177 -----VHSGKLSSEGIQYV--GOS-----VAVSENGENVL----- 205
DB 235 PHGDIADGVAVGINRLDAPRQYGIKKPSEGEYFLAEGSELKPVKAVGNNGVLIINQL 294
QY 206 -----ARLDVYRTTGGAFAANGKRLAPEPAQTDKQ 236
DB 295 NRWYKERIEONQPLSABEEMVREAOARHENLSAQAYTGG-----RLIPDP-----M 342
QY 237 DSKLSVNTVNDMQL-----DDKONINLALTHYNDKODTDYAPDYGNLAVLFG 285
DB 343 DYRGSGLAKIGYRFGGRHYGVGFEDTKQAYDIRDMTEKQYGTDEKKKFRDSSGYVDG 402
QY 286 E-----KPSLNAIKGLSLSSEQPKTKSTFNINHHDDLWGNTINTNAYR 330
DB 403 DDFPDGLYFVNIEBKMDQKLIRIGIL---KYSRTKF-IDEEH-----RR 444
QY 331 REKGRFYYPF 339
DB 445 RRMGLLYRY 453

```

```

RESULT 36
PCT-US95-06994-7
Sequence 7, Application PC/TUS9506994

```

```

; GENERAL INFORMATION:
; APPLICANT: Children's Hospital & Medical Center
; APPLICANT: University of Washington
; APPLICANT: Washington State University Research Foundation
; APPLICANT: TARR, Philip I
; APPLICANT: BILGE, SIMA S
; APPLICANT: BESSER, THOMAS E
; APPLICANT: VARY JR, JAMES C
; TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
; STREET: SUITE 2800, 1420 FIFTH AVENUE
; CITY: SEATTLE
; STATE: WASHINGTON
; COUNTRY: USA
; ZIP: WA 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: 07-JUN-95
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BRODERICK, THOMAS F
; REGISTRATION NUMBER: 31,332
; REFERENCE/DOCKET NUMBER: CHOR-18591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682 8100
; TELEFAX: (206) 224 0779
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 718 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; DESCRIPTION: Vibrio cholerae IrgA amino acid sequence,
; DESCRIPTION: wherein Xaa residues represent gaps
; DESCRIPTION: introduced to facilitate best alignment with
; DESCRIPTION: SEQ ID NO:6.
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: VIBRIO CHOLERAE
; PCT-US95-06994-7

```

```

Query Match      6.1%; Score 124.5; DB 5; Length 718;
Best Local Similarity 25.4%; Pred. No. 0.0036;
Matches 63; Conservative 35; Mismatches 95; Indels 55; Gaps 12;

```

```

QY 37 EPVITIDSGMALANRITOMPHTTVIIEEOIQEATSGROLADYMAQLPSLGVSSGT 96
DB 35 ETWVVT-----AAGYAOYIOAPASISVISREDE-----SRYRDYTDALKSVPTVXX 85
QY 97 TSFNGQT---MGRVOVQFLANGVPLTGSIDISROL--NSINP-----NOVARIE 140
DB 86 XTGGDPTDISIGMSNYTL--ILVDGKQTSRQTRRPNSDGIBQGLPPLQATERIE 143
QY 141 VLSGATS-IYSGATGGLINIVTKSLDEEOPETRIGVHSGKLSSEGIQYVGSGVAVS 199
DB 144 VIRGPMSTLYGSDALIGVNIITRKD--QQQMSGNV-----QLSTVQ 184
QY 200 ENGVNLARLDVRYRTGGARD-ANGKRIAPEAQTOD-----SKSL-SVNTVNDMQL 251

```

Db 185 ENRASGDSQSNFFVYTGPLSDXALSQVYGOTTQORDEDEIEHGYGDKSLRSILTSKLNQVL 244  
Qy 252 DDKQNINL 259  
Db 245 NPDHQLQL 252

RESULT 37  
US-08-425-843-2  
; Sequence 2, Application US/08425843  
; Patent No. 6020154  
; GENERAL INFORMATION:  
; APPLICANT: Hansen, Eric J.  
; APPLICANT: Cope, Leslie D.  
; APPLICANT: Jarosik, Gregory P.  
; APPLICANT: Hanson, Mark S.  
; TITLE OF INVENTION: H. Influenzae HxuB and HxuC Genes, Proteins  
; TITLE OF INVENTION: and Methods of Use  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/425,843  
; FILING DATE: Concurrently Herewith  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parker, David L.  
; REGISTRATION NUMBER: 32,165  
; REFERENCE/DOCKET NUMBER: AMCY:012/PAR  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (713) 789-2679  
; TELEX: 79-0924  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 725 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-425-843-2

Query Match 6.0%; Score 123.5; DB 3; Length 725;  
Best Local Similarity 23.6%; Pred. No. 0.0046;  
Matches 74; Conservative 48; Mismatches 117; Indels 75; Gaps 15;

Qy 7 FOWLSPLLSVAVTQOLYAQNPSLPTVELEP--VVITDKSGMALANRITOMPHTTKVI 64  
Db 3 FSKLSLAITLVTANALAQ-----SVELDSINVIATRDPSPAYT----- 43

Qy 65 YEEQIQEQATGSRQLADVMAQL--IPSLGVSSGTTG-----NFGQTMGRQVQFLLNGVP 117  
Db 44 -PEKOSKDSLLSKQATSVAAALIEDIPNDIRGSSRSIAQKNIRGLSDNRVVVQ-VIDGV- 100

Qy 118 LTGSRDLSRQLNSINP-NOVARIEVLVG-ATSIYSGGATGGLINVTKSLDLEEEQFETRI 175  
Db 101 -RQNFELAHRGYFIPMSLIQIEIWKIPSSSLWSSGALGGVAVNRTPNALDLLKNNKPF 159

Qy 176 GVHGSKLSSSEGIGYQVGOSVAGSNGVNLARLDVYRTTGAFDA--NGKRIAPEPAQT 233  
Db 160 GV---KIRQ---GYO-----TANNLSRDSVSFAANDKFDVLISGFFYNNADNLRT 203

Qy 234 DKQDS-----KSLSVNTNVQDLDDKQNLNIALTHYNDKQTDYAPDYGNRVLVFGKE 287

Db 204 GKGNKLNATYKQFGLGAKFGWQINDANRVELSHRETRFKQTA----- 246  
Qy 288 PSLNAIKGLSLSEQ 301  
Db 247 PSNNEVENELTNEQ 260

RESULT 38  
US-08-537-361E-4  
; Sequence 4, Application US/08537361E  
; Patent No. 6121037  
; GENERAL INFORMATION:  
; APPLICANT: Stojiljkovic, Igor  
; APPLICANT: So, Magdalene  
; APPLICANT: Hwa, Vivian  
; APPLICANT: Heffron, Fred  
; APPLICANT: Nassif, Xavier  
; TITLE OF INVENTION: No. 6121037el Bacterial Hemoglobin Receptor  
; TITLE OF INVENTION: Genes and Uses  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: McDonnell Boehrnen Hulbert & Berghoff  
; STREET: 300 South Wacker Drive, 32nd Floor  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/537,361E  
; FILING DATE: 02-OCT-1995  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 6121037nan, Kevin E  
; REGISTRATION NUMBER: 35,303  
; REFERENCE/DOCKET NUMBER: 94,784-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-913-0001  
; TELEFAX: 312-913-0002  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 791 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-537-361E-4

Query Match 6.0%; Score 122; DB 3; Length 791;  
Best Local Similarity 22.4%; Pred. No. 0.0074;  
Matches 89; Conservative 56; Mismatches 145; Indels 108; Gaps 21;

Qy 10 LSLPLLSVAVTQOLYAQP-----NESLPT-VELEPVTITDKSGMALANRITOMPHT 60  
Db 4 LQMPPIA-ALLGSIFGNPVFAADEAATETTPKAEVAVRVKQGNAPAAVERV----- 57

Qy 61 TKVIYEEQIQEQATGSRQLADVMAQLIPSLGVSSGTTGTFNFG---QTMHGRQVQFLLNGVP 117  
Db 58 LNRKIQEMIRNDKDLVRYSTDV-----GLSDRSRHRQKGFARVGEVDRGVGSIDGVN 109

Qy 118 LTGSRDTS--RQLN-----SINPNQVARIEVLVGATSI-YGSGATGGLINIVTKSD-- 165  
Db 110 LPDSEENSLYARYGNFNSSRLSIDPELVNRNIDIVKGAOSFNTGALGGGVNYQTLOQRD 169

Qy 166 --LREEQP-----ETR-----IGVHGSK-----LSSEGIGYQV-----GQSV 195  
Db 170 LLLPERQFGVMKNGYSTNRNWTNLGFGVSNDRVDVAALLYSQRRGHETESAGKRGYPV 229

Qy 196 AGVSENGVNLARLDVYRTTGTGAFDANGKRIAPEQATDKQDSKLSLVNTNVQDLDDKQ 255



Db 230 EGAGSGANI-----RGS-----RGIDPPSQ-----KHYSFLGKIAVQINDNH 268  
Qy 256 NINLALTHYNDKQDTPYADYG-NRLAVLFGEKPSLNAIKGSL-----SEOPKTKS 307  
Db 269 RIGASL---NGQGHNTYTESYNLLASYREADVNRNRNTNLFYEWTPESDRLSMVKA 325  
Qy 308 TFN-----INHHDDLWGNTINTNAYRRKGRFY 337  
Db 326 DVDYOKTKVSAVNYKGSFPTNYTTWETETHEHKEVEEYI 363

RESULT 39  
US-08-817-707-4  
; Sequence 4, Application US/08817707  
; Patent No. 6277382  
; GENERAL INFORMATION:  
; APPLICANT: Stojiljkovic, Igor  
; APPLICANT: So, Magdalene  
; APPLICANT: Hwa, Vivian  
; APPLICANT: Heffron, Fred  
; APPLICANT: Nassif, Xavier  
; TITLE OF INVENTION: No. 6277382el Bacterial Hemoglobin Receptor  
; TITLE OF INVENTION: Genes and Uses  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
; STREET: 300 South Wacker Drive, 32nd Floor  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/817,707  
; FILING DATE: 19-AUG-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 6277382nan, Kevin E  
; REGISTRATION NUMBER: 35,303  
; REFERENCE/DOCKET NUMBER: 94,784-J  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-913-0001  
; TELEFAX: 312-913-0002  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 791 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-817-707-4

Query Match 6.0%; Score 122; DB 4; Length 791;  
Best Local Similarity 22.4%; Pred. No. 0.0074;  
Matches 89; Conservative 56; Mismatches 145; Indels 108; Gaps 21;

Qy 10 LSLPLLSAVNTQOLYAOP-----NESLPT-VELBPVYITIDKSGMALNRIITOMPHT 60  
Db 4 LQMPPIA-ALIGSIFGNVFAADEATETTPVKAIVKAVKGGRNAPAAVERV-----N 57  
Qy 61 TKYIEEOIQEATGSQLADYMAQLIPSLGVSSGTTSNNG---QTHAGROVQPLANGVP 117  
Db 58 LMRKQEMIRDNKDLVRYSTDV-----GLSDRSRHQKGFALRGVGGVGVGVSIDGVN 109  
Qy 118 LTGSRDIS---ROLN-----SINPNQVARIIEVLGATSI-YGSGATGGLINITYTKSD-- 165  
Db 110 LPPSEENSLVARYGNPSSRLSIDPELVNRNDIYKGDSPRTSGALGGGVNYVTLOGRD 169  
Qy 166 --LEEQF-----ETR-----IGVHSGK-----LSSEGLGYQV-----GQSV 195

Db 170 LLLPERQFVMMKNGYSTENREMTNTLGFVGNDRDALLYSORGHETESAGKRGYPV 229  
Qy 196 AGVSENGVNLARLDVDRYRTGGAFPANGKRIAPEPAQDTGKQSKLSVNTVNDOLDLQ 255  
Db 230 EGAGSGANI-----RGS-----RGIDPPSQ-----KHYSFLGKIAVQINDNH 268  
Qy 256 NINLALTHYNDKQDTPYADYG-NRLAVLFGEKPSLNAIKGSL-----SEOPKTKS 307  
Db 269 RIGASL---NGQGHNTYTESYNLLASYREADVNRNRNTNLFYEWTPESDRLSMVKA 325  
Qy 308 TFN-----INHHDDLWGNTINTNAYRRKGRFY 337  
Db 326 DVDYOKTKVSAVNYKGSFPTNYTTWETETHEHKEVEEYI 363

RESULT 40  
US-08-817-707-6  
; Sequence 6, Application US/08817707  
; Patent No. 6277382  
; GENERAL INFORMATION:  
; APPLICANT: Stojiljkovic, Igor  
; APPLICANT: So, Magdalene  
; APPLICANT: Hwa, Vivian  
; APPLICANT: Heffron, Fred  
; APPLICANT: Nassif, Xavier  
; TITLE OF INVENTION: No. 6277382el Bacterial Hemoglobin Receptor  
; TITLE OF INVENTION: Genes and Uses  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff  
; STREET: 300 South Wacker Drive, 32nd Floor  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/817,707  
; FILING DATE: 19-AUG-1997  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 6277382nan, Kevin E  
; REGISTRATION NUMBER: 35,303  
; REFERENCE/DOCKET NUMBER: 94,784-J  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-913-0001  
; TELEFAX: 312-913-0002  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 790 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-817-707-6

Query Match 5.9%; Score 121; DB 4; Length 790;  
Best Local Similarity 23.1%; Pred. No. 0.0092;  
Matches 93; Conservative 50; Mismatches 136; Indels 124; Gaps 24;

Qy 12 LPLLSAVNTQOLYAOP-----NESLPT-VELBPVYITIDKSGMALNRIITOMPHTTK 62  
Db 7 LPI--AALVGSIFGNVFLADEATETTPVKAIVKAVKGGRNAPAAVERV-----NLN 59  
Qy 63 VYIEEOIQEATGSQLADYMAQLIPSLGVSSGTTSNFG---QTHAGROVQPLANGVPLT 119  
Db 60 RIKQEMIRDNKDLVRYSTDV-----GLSDSGRHQKGFAPVAVRGVGGVGVGVSIDGVN 111  
Qy 120 GSRDIS---ROLN-----SINPNQVARIIEVLGATSI-YGSGATGGLINITYTKSD--- 165

Db 112 DSEENSLVARYGNFNSRLSIDPELVRIEIVKGADSFNTSGALGGVNYQTLOGRDL 171  
Qy 166 LEEQF-----ETR-----IGVHGSK-----LSSEGIGYQV-----QOSVAG 197  
Db 172 LDDROFGVMKNGYSTRWNTLIGFVSNDRVDAALLYSORRHETESAGNRCYPVEG 231  
Qy 198 VSENGNVLARLDVYRTTGGAFDANGKRIAPAPQTDKQDSKLSVNTNVDKLDKONI 257  
Db 232 AGKETNI-----RGSA-----RGIPDPSK-----HKYHFLGKIAYQINDNHR 270  
Qy 258 NLALHYNDKQTDY-----APDYNRL-AVLPGF-KESLNAIKGLSL- 299  
Db 271 GASL---NGQOHNTVVEESYNLTASSWEADVVNRNRNANLFYEMWPESNWLSSLKADF 327  
Qy 300 EOPKT-----TKSTFNINVYHDDLWNTINTNAYYRREKGRFY 337  
Db 328 DYOKTKVAANKGSPTNY-----TTWETEHKKEVGEIY 362

RESULT 41  
US-08-363-124A-4  
; Sequence 4, Application US/08363124A  
; Patent No. 5912336  
; GENERAL INFORMATION:  
; APPLICANT: Sparling, P. Frederick  
; APPLICANT: Cornelissen, Cynthia N.  
; TITLE OF INVENTION: Transferrin-Binding Proteins From  
; TITLE OF INVENTION: Neisseria Gonorrhea and Neisseria Meningitidis  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: ImClone Systems Incorporated  
; STREET: 180 Varick Street  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10014  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/363.124A  
; FILING DATE: 23-DEC-1994  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/124,254  
; FILING DATE: 20-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/973,336  
; FILING DATE: 05-NOV-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/572,187  
; FILING DATE: 23-AUG-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gallagher, Thomas C.  
; REGISTRATION NUMBER: 37,066  
; REFERENCE/DOCKET NUMBER: SPA-1-PDC  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 645-1405  
; TELEFAX: (212) 645-2054  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 909 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-363-124A-4

Query Match 5.9%; Score 121; DB 2; Length 909;  
Best Local Similarity 20.6%; Pred No. 0.012;  
Matches 89; Conservative 60; Mismatches 161; Indels 122; Gaps 23;

Qy 1 MRHHYFQWLSPLLSVAVTQQLYAO--PNESLPVLEBPVVTIDKSGMALANRITOMP 58  
Db 1 MOOHLFR-LNILCSLMTALPVYAENVQAEQAKOLDTIQKAKKQKTRRDNEVTGLG 59  
Qy 59 HTTKVIYEEQIQEATGSRDLADVMAQLIPSLG-VSSGTTSNFGQTMHG---ROVQFLN 114  
Db 60 KLVK-SSDTLSKEQVLNIRD----TRYDPIAVVEQGRGASSGYSIRGMDKNRVSITVD 114  
Qy 115 GV-----PLTGSRDI--SRQLNSINPNQVARIEVLSGA-TSIYGSATGGLINIV 161  
Db 115 GVSQIQSVTAQAALGGTRTAGSSGAINIEYENVKAVEISKGSNSSEYGNALAGSVAPO 174  
Qy 162 TK--SDL--BEEQPETRIGVHSGKLSSEGIGYQVQSGVAGVSENGNVLAR----- 207  
Db 175 TKTAADIIGEGKQW---GIQ-SKTAYSGKDHALTQSLALAGRSGGAELLIYTKRRGRE 229  
Qy 208 -----LQVDYRTTGGAFDA-----NGKRIAPEACTD---KODS 238  
Db 230 IHAHKDAGKGVQSPNRLVLDDEKKEGGSQSDISLCEECHNGYAAACNKKUKEDASVKDER 289  
Qy 239 KLSLVNT-----NVDWQLDDKQINLALTHYNDKQDT----- 270  
Db 290 KTVSTQDYTGSNRLANPLEYGSQSQSWLFRPGWHLNDRHYGAVLERQOTQDTRDMTVP 349  
Qy 271 -----DYAP-----DYGNRVLAVLF--GEKPSLNAI-KGLSLSEOPKTKTKTFNINY 313  
Db 350 YFTSEYVPGSLKGLKGYSGDNKAERLFVQEGSTLOGIGYGTGTFYDERHTKRYGVEY 409  
Qy 314 -HHD---DLGN 321  
Db 410 VYHNADKDTWAD 421

RESULT 42  
US-09-074-658-75  
; Sequence 75, Application US/09074658  
; Patent No. 6184371  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena M  
; APPLICANT: Run-Pan Du  
; APPLICANT: Quijun Wang  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: LACTOFERRIN RECEPTOR GENES OF MORAXELLA  
; NUMBER OF SEQUENCES: 78  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/074,658  
; FILING DATE: 08-MAY-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-795  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 75:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 941 amino acids  
; TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-074-658-75

Query Match  
Best Local Similarity 19.7%; Pred. No. 0.021;  
Matches 96; Conservative 56; Mismatches 148; Indels 187; Gaps 21;

QY 1 MRSHYFQWLSPLSLAVATQOQYAPNE-SLPTVELEPVVITIDKSGMALANRITQMPH 59  
DB 1 MNKHSFP-LTLTALAIATAPPSAANSEIATQOSLKEVTAATKAG-----RSKXEVTS 55  
QY 60 TTYVYIEEQI--QEOATGSRQADVAQOLIPSLG-VSSGTTNFGQTWMC--RQVQFLN 113  
DB 56 LGKIVKTSKTEKQVIGIRDL-----TRYDPGVAVVEQNGAGSGYSIRGVDKRVAVSV 111  
QY 114 NGVPLT-----GSRDISRQINSINPQVARIETVLSGA-TSIVSGATGLINI 160  
DB 112 DGAQIOAFTVQSLSGYGGSGGAINELIYENISTVEIDKGGSSDHSGALGAVAF 171  
QY 161 VTK--SDL-----EEOFEETRIGVHSGKLSSEGI-----187  
DB 172 RTEADLDISDGSKGIGQATKATGSKRQKTMKSLGAGFSKDGHEGLIRTERQRETRPH 231  
QY 188 ---GYVG-----QSVAGVSENGVNL-----205  
DB 232 GDIADEVEYIDRLDAFRQTYDIQKONKKAIFYLAEGESELIKPAKLAGNGVTLKQNLNR 291  
QY 206 -----ARLDVYRTTGAFDANGKRIAEPAQTDKQDS 238  
DB 292 WBERKKNOSLSAEEMVREAOARHENISAQVYGG-----RILDP-----MDY 339  
QY 239 KSLSVNTVDMOL-----DDKONINLALTHYNDQDYPADYGNRLAVLFG- 286  
DB 340 RSSSWAKLGYRRGGRRHYGVGVEDTKQRDIDMTKQYTGIDEPKFDKGYNDGD 399  
QY 287 -----KESLNAIKGLISLSEQPTKSTFNINVHDDLMGNTINTNAVYRRE 332  
DB 400 FRDLGYFVPIEEMKGDKNLVKIGL-----KYSRTKE-IDENH-----RRRR 441  
QY 333 KGRFYFP 339  
DB 442 MGLLYRY 448

RESULT 43  
US-08-487-890A-94  
Sequence 94, Application US/08487890A  
Patent No. 5708149  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena  
APPLICANT: Harkness, Robin  
APPLICANT: Schryvers, Anthony  
APPLICANT: Chong, Pele  
APPLICANT: Gray-Owen, Scott  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Mordin, Andrew  
APPLICANT: Klein, Michel  
TITLE OF INVENTION: Transferrin Receptor Genes  
NUMBER OF SEQUENCES: 147  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1K7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/487, 890A  
FILING DATE: 07-JUN-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,116  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/148,968  
FILING DATE: 08-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-466 MTS-jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 94:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 908 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-487-890A-94

Query Match  
Best Local Similarity 20.6%; Pred. No. 0.022;  
Matches 89; Conservative 59; Mismatches 162; Indels 122; Gaps 22;

QY 1 MRSHYFQWLSPLSLAVATQOQYAPNE-SLPTVELEPVVITIDKSGMALANRITQMPH 58  
DB 1 MOOHLFR-LNLICSLIMTALPYAENVQAEQAEKQDLTIQVAKKQKTRRDNEVTGIG 59  
QY 59 HTTKVYIEEQI--QEOATGSRQADVAQOLIPSLG-VSSGTTNFGQTWMC--RQVQFLN 114  
DB 60 KLVK-SSDLSKQVUNITDL-----TRYDPGIAVVEQNGAGSGYSIRGMDKRVSLTYD 114  
QY 115 GV-----PLTGSRI--SRQINSINPQVARIETVLSGA-TSIVSGATGLINI 161  
DB 115 GVGQISYTAQALGSTRAGSSGAINELIYENIVKAVEISKNSSEYNGALAGSVAQ 174  
QY 162 TK--SDL--EEOFEETRIGVHSGKLSSEGIQVGGSVAGVSENGVNLAR-----207  
DB 175 TKYAADIIIGEGKOW-----GIO-SKTAYSGKDHALTQSLAGSGAGALLIYTKRGRE 229  
QY 208 -----LDVDYRTTGAF-----DANKRIA 227  
DB 230 IHAHKDAGKGVOSFNLVLDEDEKGGQRYFYIEBECNGYACKNLKEDASVQDER 289  
QY 228 PEPAQTDKQDSKLSVNT-----NVDMQDDKONINLALTHYNDKQD-----270  
DB 290 KTVSTQDYTGSRNRLANPLEYGSQSWLFRPGHLDNRHYGAVLERTQOTFTDRMTVA 349  
QY 271 ----DYAP-----DYGRLAVLF--GEKPSINAL-KGLISLSEQPTKSTFNIN 313  
DB 350 YFTSEDYVVGSLKGLKYGSGDNKAERLFGVGGSTLGGIGYGVYDRAHRTKNRGVY 409  
QY 314 -HHD---DLMGN 321  
DB 410 VYHNAKDXTWAD 421

RESULT 44  
US-08-478-435-94  
Sequence 94, Application US/08478435  
Patent No. 5922323  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena  
APPLICANT: Harkness, Robin  
APPLICANT: Schryvers, Anthony  
APPLICANT: Chong, Pele  
APPLICANT: Gray-Owen, Scott  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Mordin, Andrew



Qy 208 -----LDVYRTTGAF-----DANGRIA 227  
Db 230 IHAHKDAGKGVOSFNNRLVLDEDEKKGGSQYRYFIVEECHNGYAACNKLKEDASVQDER 289  
Qy 228 PEPAQTDKQDSKLSVNT-----NVDWQDDKONINLALTHYNDKQDT-----270  
Db 290 KTVSTQDYTGSSNRLNPLEYGSQSWLFRPGWHLDNHRYVGAVLERTQOTFDTDRMTVPA 349  
Qy 271 ----DYAP-----DYGNNRLAVLF--GEKPSLNAI-KGLSLSEQPKTKSTFNINY 313  
Db 350 YFTSEDIYVPGSLKGLGKYSGDNKAERLFPVQEGSITLOGIGYGVFYDERHTKNRYGVEY 409  
Qy 314 -HHD---DLMGN 321  
Db 410 VYHNADKDTWAD 421

Search completed: December 25, 2002, 20:22:51  
Job time : 38 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compen Ltd.

OM protein - protein search, using sw model

Run on: December 25, 2002, 20:17:03 ; Search time 42 Seconds

(without alignments)  
915.566 Million cell updates/sec

Title: US-09-889-746-2

Perfect score: 2047  
Sequence: 1 MHSHYFQMLSLPLSLVAVT.....SKAEVLGRVPLNKPRLALF 400

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	527	25.7	813	2 P83476	probable siderophore
2	503.5	24.6	742	2 C83061	probable TonB-depe
3	491	24.0	726	2 AB0122	probable ferric B1
4	472	23.1	732	2 T44483	receptor-like prot
5	457	22.3	725	2 S01042	aerobactin recepto
6	453.5	22.2	851	2 S74442	ferric aerobactin
7	453	22.1	746	2 T46821	siderophore recept
8	453	22.1	746	2 A95420	RhA Rhizobactin r
9	400.5	19.6	867	2 AD1856	hypothetical prote
10	333.5	16.3	536	2 AF2128	hypothetical prote
11	326.5	16.0	536	2 AE2133	ferric aerobactin
12	326.5	16.0	536	2 AC2082	ferric aerobactin
13	200	9.8	726	2 AG0837	TonB-dependent out
14	187.5	9.2	726	2 AG2219	hypothetical prote
15	187	9.1	867	2 AC2140	ferrichrome-iron r
16	182	8.9	858	2 AC2134	ferrichrome-iron r
17	181.5	8.9	858	2 AD2421	ferrichrome-iron r
18	180	8.8	746	2 AC2079	ferrichrome-iron r
19	179.5	8.8	746	2 A40636	ferric enterobacti
20	178.5	8.7	802	2 AC2075	outer membrane hem
21	172	8.4	747	1 QRECFE	ferrichrome-iron r
22	171	8.4	747	2 B85499	outer membrane rec
23	168.5	8.2	696	2 B81346	probable iron upre
24	168.5	8.2	696	2 DB1357	hemelin uptake syst
25	168.5	8.2	820	2 AE2130	ferrichrome-iron r
26	168.5	8.2	820	2 H83529	siderophore recept
27	165.5	8.1	742	2 AI2077	ferrichrome-iron r
28	165.5	8.1	863	2 AI2077	ferrichrome-iron r
29	163.5	8.0	849	2 AI2129	ferrichrome-iron r

30	159.5	7.8	863	2 S74447	ferrichrome-iron r
31	159.5	7.8	883	2 C83385	hypothetical prote
32	159	7.8	878	2 AH2075	ferrichrome-iron r
33	155.5	7.6	706	2 AH0670	probable TonB-depe
34	155	7.6	788	2 C82595	ferric enterobacti
35	154.5	7.5	660	2 D91176	heme utilization/c
36	154.5	7.5	660	2 E86022	outer membrane hem
37	154.5	7.5	746	2 G90706	hypothetical prote
38	154.5	7.5	746	2 B85557	hypothetical prote
39	153	7.5	718	2 A83282	probable TonB-depe
40	153	7.5	817	2 D87390	TonB-dependent rec
41	152	7.4	710	2 C96235	probable hydroxama
42	152	7.4	716	2 A13050	hydroxamate-type f
43	151.5	7.4	611	2 A82358	vitamin B12 recept
44	151.5	7.4	676	2 T12069	hemin receptor - y
45	151.5	7.4	676	2 AG0035	hemin receptor pre

## ALIGNMENTS

## RESULT 1

F83476 Probable siderophore receptor PA1365 [imported] - Pseudomonas aeruginosa (strain PA01)  
C/Species: Pseudomonas aeruginosa  
C/Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C/Accession: F83476  
R/stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Lattig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A/Reference number: A82950; PMID:20437337; PMID:10964043  
A/Accession: F83476  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-813 <STO>  
A/Cross-references: GB:AE004565; GB:AE004091; NID:g9947294; PIDN:AA04754.1; GSPDB:GN001  
A/Experimental source: strain PA01  
C/Genetics:  
A/dene: PA1365

Query Match	Score	25.7%	DB 2;	Length	813;
Best Local Similarity	34.2%	Pred. No. 2.5e-29;			
Matches 137;	Conservative	68;	Mismatches 144;	Indels	52;
				Gaps	9;
QY 2 RSHSYFQMLSLPLSLVAV-TQQLYAQPNSLPTVELEPVVITDKSGMALANRITQMPHT 60					
DB 94 RDEHTE--IVVPASEAAVPATQARSEP-----LDWEQMEITFASRTSSDLVSATRO---- 141					
QY 61 TKTYIEQIQEQATGSRQADVMAQLIPSLGVSSGTTSNFGQTMHQRQVQFLNGVPLTG 120					
DB 142 STVEHAQLEELRQSSDSLATVLAQVPMQSDSRITTYEGQTLRGRSMVWVDVGPLNT 201					
QY 121 SRDISRLNSINPNQVARIIVLSGATGATGALINIVTKSDLEBEQFETRIGVHG- 179					
DB 202 NRSSNNLANIDPALIERIEVIRGSSAITYGSGATGIIITTRPAGENRAERLRLATSP 261					
QY 180 -SKLSEGGIGYQGVGSAVGSBENGVLARLDVY---RTTGAFDANGKRLAPSPAQTD 234					
DB 262 LTRLGSDGLGGQFOQYFAG-----SLGALDYSFDFGTRHVGASDAHGRIAPBPSQGD 315					
QY 235 KQDSKSLSVNTVNDQDDKQINILATLHYNDKQDIPADYGNRLAVLFGEKPSLNATK 294					
DB 316 LFFSNVYNIGGKGLRIDENQVQLALSHDARQDDVDYATD--PRVAARLPQGSVPAAR 373					
QY 295 GLLSSEQPKTKTKTFNINHYHDDLMGNTINTNAYYRREKGRFYFVAPPSIAKALPILQS 354					
DB 374 GLLDQDNRIKRLTLAELENDILGSRLSAQLYYDYDFTRFRFPDA----- 420					
QY 355 WNLPSATLQATYAPARAAGVGLQSSSKAEVLGRVPLNKP 395					
DB 421 -----RAVSTRGGNVQIMQNSVFGSRLLTATP 449					

```
RESULT 2
C83061
probable TonB-dependent receptor PA4675 [imported] - Pseudomonas aeruginosa (strain PA01
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: C83061
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: AB2950; MUID:2043737; PMID:10984043
A:Accession: C83061
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-742 <STO>
A:Cross-references: GB:AE004882; GB:AE004091; NID:9950939; PIDN:AA08062.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4675

Query Match 24.6%; Score 503.5; DB 2; Length 742;
Best Local Similarity 30.8%; Pred. No. 1e-27;
Matches 132; Conservative 80; Mismatches 171; Indels 45; Gaps 10;

Qy 11 SLPLLSVAVTQQLYAQPNSLPTVELEPVVITIDKSGMALANRITOMPHTTKVIEEQIQ 70
Db 4 SIPLRPALSLSLFASFSAPALADPEQOMVIGSRAPTSISELPCTVWVIERQLD 63

Qy 71 EQATSGRLADVMAQLIPSLGVSSGTTNFQTMHGRQVQLNGVPLTGRDISRQLNS 130
Db 64 QQTQAGVPLKEALGQLIPGLDITGSGQRTNNGQNLGRSVLVMIDGVLNLSRGISRPQDS 123

Qy 131 INPNOVARIEVLGATSIYSGATGGLINIVTKSDL-EEQFETRIGVHGSKLSEGIY 189
Db 124 IDPFNIEIRVNSGASAVYGGATGGIINIVTKGVGGDTFRNTELGARSGQSHEDHL 183

Qy 190 QVGSGVAGVSENGVNL--ARLDVDYRTTGGAFDANGKRIAPPAQTDKQDSKLSVNTNV 247
Db 194 RAAQSISG---GNDLFNGRLAIAVQKNGAYDGGDQVLTDTITDQLQYNSVDLMGSL 239

Qy 248 DWQLDDKQINIALTHYNDKQDTPYAPYGNRLAVLFGEKPSLNAIK-GLSLSEQPTTK 306
Db 240 GFTFANGHSLDLGLOYSGYDGRGLDGRNFDALGRAP--YSIKGGVDLDRPESEKR 297

Qy 307 STEFNIYHDDLWGNTINTNAYYREKGRFVYPF-----VAPPSIAK----- 347
Db 298 HOFNATYHAPVLGHDLVQLQAYYRKEAFNFPPTIRYNSNTGAINGYTSYASQDDTDY 357

Qy 348 ---ALPILQSNLPSATL-----DAYT-----KAPQARAYGVLSQSKAEVLGRVNL 392
Db 358 YGMKLALVKTWERASLYGVLDREKFTSDQMLFNLPLAASGGLVASEQAK-LGRYDI 416

Qy 393 NKPKRALP 400
Db 417 DTDSRAFF 424

RESULT 3
AB0122
probable ferric siderophore receptor iutA [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AB0122
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Fitball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AB0122
```

```
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-726 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC9837.1; PID:gl5979062; GSPDB:GN00175
C:Genetics:
A:Gene: iutA

Query Match 24.0%; Score 491; DB 2; Length 726;
Best Local Similarity 30.7%; Pred. No. 7.5e-27;
Matches 123; Conservative 69; Mismatches 155; Indels 54; Gaps 9;

Qy 1 MRHHYFQWLSPLLSVAVTQQLYAQPNSLPTVELEPVVITIDKSGMALANRITOMPHT 60
Db 1 MKHKLH--WVLNPCLLVWLTPAAWAE-----DQLVVSANRSHRSVA-----EMAQT 44

Qy 61 TKVIYEIQIQEATGSRQLADVMAQLIPSLGVSSGTTNFQTMHGRQVQLNGVPLTG 120
Db 45 TWVIEGQLEQVQGGLEIKDILAQLPIDVSSQRTNYGMNMRGHSIMVMIDGVLNS 104

Qy 121 SRDISRQLNSINPNOVARIEVLGATSIYSGATGGLINIVTKSDLEEEQFETRIGVHGS 180
Db 105 SRSDSRQLDSIDPFNIAHIEIVISGATSLYGGSGTGLINIVTKKGQEQVQLQIGKGTG 164

Qy 181 KLSSEGIGYQVGSVAGVSENGVNLARLDVDYRTTGGAFDANGKRIAPPAQTDKQDSKS 240
Db 165 FNSHNDHDENISAAMSGTE--RAFGRFSVSYQRYGMYDGKNEVLIDNTQTGLQYSNR 222

Qy 241 LSVNTNVDMQLDDKQINIALTHYNDKQDTPYAPYGNRLAVLFGEKPSLNAIKGLSLSE 300
Db 223 LDVMTGTINIDENQQLTQYFNSDSGKHGLYLGQNSAVTGTGQASNS--AALNSDR 281

Qy 301 QPKTKSTFNINYHDDLWGNTINTNAYYREKGRFPYFPVAPPSIAKALPILQ----- 353
Db 282 IPGTERHLINQYNSNTDFWQDLVAQVYRDESLTFYYPF-----PTLKQGVSTI 331

Qy 354 -----SMNLPSATLDAYTKAPQARAYGV-LQSES 381
Db 332 GASQKTDYFGSKLTANSEPIDSLT-----LTYGIDLEHS 367

RESULT 4
T44483
receptor-like protein iutA [imported] - Shigella flexneri
C:Species: Shigella flexneri
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T44483
R:Moss, J.E.; Cardozo, T.J.; Zychlinsky, A.; Groisman, E.A.
Mol. Microbiol. 33, 74-83, 1999
A:Title: The selC-associated SHI-2 pathogenicity island of Shigella flexneri.
A:Reference number: 222779; MUID:99340540; PMID:10411725
A:Accession: T44483
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-732 <MOS>
A:Cross-references: EMBL:AF141323; NID:95532445; PIDN:AAD44750.1; PID:95532466
A:Experimental source: strain M90T; serotype 5a
C:Genetics:
A:Gene: iutA

Query Match 23.1%; Score 472; DB 2; Length 732;
Best Local Similarity 30.9%; Pred. No. 1.7e-25;
Matches 130; Conservative 65; Mismatches 176; Indels 50; Gaps 9;

Qy 10 LSLPLLSVAVTQQLYAQPNSLPTVELEPVVITIDKSGMALANRITOMPHTTKVIEEQI 69
Db 14 LLLVVLSPVAQQ--NDNE-----IIVSASRNTVA----EMAQTWTWIENAEI 58

Qy 70 QEQTATGSRQLADVMAQLIPSLGVSSGTTNFQTMHGRQVQLNGVPLTGRDISRQLN 129
Db 59 BQIQGGKELKDALAQLPGLDVSSQRTNYGMNMRGRLVLLIDGVLNLSRSDSRQLD 118

Qy 130 SINPNOVARIEVLGATSIYSGATGGLINIVTKSDLEEEQFETRIGVHGSKLSEGIY 189
Db 131 SINPNOVARIEVLGATSIYSGATGGLINIVTKSDLEEEQFETRIGVHGSKLSEGIY 189
```



Db 119 SVDPFNIDHIEVTSAGATAYGGSTGCLINIYTKGQPEFMEEFAGTKGFGNSSKDHE 178  
 Qy 190 QVGSVAVSENGENVLRLDVEDYRTTGAFDANGKRIAPPAOTKODKSKLSVNTVNW 249  
 Db 179 RINGAVSG--GNHISGRLSVAAYQKFGWFDGNDATLLNDTQGLSHNRDLDMGTGL 236  
 Qy 250 QLDDKQINIALTHYNDKODTDYADYGNRLAVLFGK-PGLNAIKGLSLEOPRTTKST 308  
 Db 237 NIDESRQLQITQYKSGQGDNYGLNGKGFSAISGSGSTPEVS--KGLNSDRIPGTERHL 294  
 Qy 309 FNNINHHDDLWGNTINNTNAYRREKGFYFPVAPFSIAKALPILOS-----MMLP 356  
 Db 295 ISLQYSDSDFLQGLVGVYVRDESLRYPPPTVANKQATFESSQODTDQYGMKLTLLN 354  
 Qy 359 SATLDA-----YTKAPQARAYGVLOSSEKAEVGRVYNLKPRL 399  
 Db 355 SQMLDQGQITWGLDAHERFTSNMFPDLQASASGLLNH-KIYTTGRIPSYDITMLAA 413  
 Qy 400 F 400  
 Db 414 F 414

## RESULT 5

S01042  
 aerobactin receptor precursor iutA [validated] - Escherichia coli plasmid ColIV-K30  
 N:Alternate names: cloacin receptor  
 C:Species: Escherichia coli  
 C>Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 26-May-2000  
 C/Accession: S01042  
 R:Krone, W.J.A.; Stegehuis, F.; Koningsstein, G.; van Doorn, C.; Roosendaal, B.; de Graaf  
 FEMS Microbiol. Lett. 26, 153-161, 1985  
 A>Title: Characterization of the pColIV-K30 encoded cloacin Df13/aerobactin outer membra  
 e sequence and primary structure.  
 A:Reference number: S01042  
 A/Accession: S01042  
 A/Molecule type: DNA  
 A:Residues: 1-725 <KRO>  
 A/Cross-references: EMBL:X05874  
 C:Genetics:  
 A:Gene: iutA  
 A:Genome: plasmid ColIV-K30  
 C:Function:  
 A:Description: functions as outer membrane receptor for ferric aerobactin [validated, M  
 F:1-25/Domain: signal sequence #status predicted <Sig>  
 F:26-725/Product: cloacin receptor #status predicted <Mat>

Query Match 22.3%; Score 457; DB 2; Length 725;

Best local Similarity 30.8%; Pred. No. 1.9e-24;  
 Matches 133; Conservative 61; Mismatches 178; Indels 60; Gaps 9;

Qy 6 YFQV----LSPLISVAVTQOLVAQPNESLPTVELEPVITIDSGMALNR-----ITOM 57  
 Db 7 YTTMALNPLLTMTAPVAVAO-----TDDRTFYVANSRNRRTYAE 47  
 Qy 58 PHTTKVIEBOIQOAGSROLADVMAQLPGLSVSGSTSNFGQTMHGRQVPLANGVP 117  
 Db 48 AQTITVIEANLEEQIGGKELKALQOLPGLDVSSRSRNYGMNVRKRLVVLVGVSR 107  
 Qy 118 LTSGRDISROLNSINNOVARIEVLASGATISYGATGGLINITYKSDLBEOPETRIGV 177  
 Db 108 LNSRSTSRQDLSIDPFNMHHIEVIFGATSLYGGSGTGGLINITYKSGOPEFMEEFAGT 167  
 Qy 178 HGSGLSEEGIGYQVGSVAVSENGENVLRLDVEDYRTTGAFDANGKRIAPPAOTKOD 237  
 Db 168 KSGFSSSKDHERIYAGVSGNE--HISGRLSVAYOKFGWFDGNDATLLNDTQGLQY 225  
 Qy 238 SKLSVNTVNWQDLDKQINIALTHYNDKODTDYADYGNRLAVLFGKESLNALIGLS 297  
 Db 226 SDRDLINGTGLINIDBSROLITQYKSGQDDYGLNLGSGFAIKGTSTPF-VSNGLN 264  
 Qy 298 LSEQPKTKSTFNINHHDDLWGNTINNTNAYRREKGFYFP-----VAPFSIAKAL 349

Db 285 SDRIPGTERHLISLQYSDSAFLQGLVGVYVRDESLRYPPPTVANKQVTAFFSSQOD 344  
 Qy 350 PILQSNMLPEAT-----LDA-----YTKAPQARAYGVLOSSEKAEVIGR 388  
 Db 345 TDQYGMKLTLLNSRPMQWQITWGLDADHERFTSNMFPDLQASASGLLNK-KIYTTGR 403  
 Qy 389 VPINLKPRLP 400  
 Db 404 YPSYDITNLA 415

## RESULT 6

S74442

ferric aerobactin receptor iutA - Synechocystis sp. (strain PCC 6803)

N:Alternate names: protein s111206

C:Species: Synechocystis sp.

A/Variety: PCC 6803 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000

C/Accession: S74442

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.  
 O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasud  
 DNA Res. 3, 109-136, 1996A>Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocysti  
 s.

A:Reference number: S74322; MUID:97061201; PMID:8905231

A/Accession: S74442

A&gt;Status: nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A:Residues: 1-851 &lt;KAN&gt;

A/Cross-references: EMBL:D90899; GB:AB001339; NID:g1651650; PIND:BA16594.1; PID:g16516

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Gene: iutA

A:Start codon: GNG

C/Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homo

Query Match 22.2%; Score 453.5; DB 2; Length 851;

Best local Similarity 33.0%; Pred. No. 4.3e-24;  
 Matches 113; Conservative 67; Mismatches 143; Indels 19; Gaps 10;

Qy 22 QLIAQPNESLPTVELE-----PVITIDSGMALNRITOMPHTKVIEBOIQOATGS 76  
 Db 153 QLIIRPVGAGTARIESEBALQIVSATRTBERIAN----IPRSVYIEBAIEEQOTQV 208  
 Qy 77 ROLADVAQOLIPGLSVSGSTSNFGQTMHGRQVPLNGVPLTSGRDISROLNSINNOV 136  
 Db 209 SSLADIVGLVPLGAPSTGSASQFGALRGRNVLVLDGVPTTNRAFPDLQTPASAI 268  
 Qy 137 ARIEVLASGATISYGATGGLINIVTKSDLBEOF--ETRIGVHGSKLS--SEGIGYQVGS 194  
 Db 269 ERIEFVIOGPTAIGDAGTGVINIIITRGPAPRPPPLASTRLAINTDPVSNSLCRMVEQY 328  
 Qy 195 VAVSENGENVLRLDVEDYRTTGAFDANGKRIAPP-AQTDKODKSKLSVNTVNWQDLD 253  
 Db 329 VGGTLVDVADY--AFTASYESVGGFALGNRIPPDPNGQGVSDTDSFNVLGKLGIMNTD 386  
 Qy 254 KQINIALTHYNDKODTDYADYGNRLAVLFGKPSLNALIKGLSLEOPRTTKST-FNIN 312  
 Db 387 EQRLOTINHPQATQNTDFTVD--PSITLALAGQRS-QAIDGLDL-DTPQTSNNTVSLD 442  
 Qy 313 YHHDDLWGNTINNTNAYRREKGFYFPVAPFSIAKALPILQS 354  
 Db 443 YSHSNVNLGNLKGQIYTRDYLTFRFPDGSASVLSGNSITQS 484

## RESULT 7

T46821

siderophore receptor iutA, tonB-dependent [imported] - Rhizobium melioli

C/Species: Rhizobium melioli

C/Date: 17-Mar-2000 #sequence\_revision 17-Mar-2000 #text\_change 17-Mar-2000

C/Accession: T46821

R:Lynch, D.; O Connell, M.; O'Brien, J.  
 submitted to the EMBL Data Library, December 1998

A;Description: Cloning and sequence analysis of the Sinorhizobium meliloti 2011 rhizobial  
A;Reference number: 224097  
A;Accession: T46821  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-746 <LYN>  
A;Cross-references: EMBL:AF110737; PIDN:AAD09419.1  
A;Experimental source: strain 2011  
C;Genetics:  
A;Gene: rhtA  
C;Function:  
A;Description: involved in ferrisiderophore uptake

Query Match 22.1%; Score 453; DB 2; Length 746;  
Best Local Similarity 32.6%; Pred. No. 3.9e-24;  
Matches 109; Conservative 67; Mismatches 132; Indels 26; Gaps 9;

Qy 18 AVTQOLYAQPNESTPVELEPVVITIDKSGMALANRITQMPHTTKVIYEEQIQEQATGSR 77  
Db 24 AVAQE---PANQSEAVTSLEIIVVTGGRS---AQQISEARTIYVVDSDQIQAEARSGK 76  
Qy 78 QLADVMAQLIPSLG-VSSGTTSNFGQTMHGQVQFLNGVPLTGSRDLSRQLNSINPNQV 136  
Db 77 TLQILGETIFSPDASDGARTSFQNLGRPLILVDGVSMSARSLSRQFDALDPENI 136  
Qy 137 ARIEVLGATSIYGGATGGLINIVTKSDLEEEQFETRIGVHGSKLSEGGIGYQVGQ--- 193  
Db 137 ERVEVLGATAIYGNATGGIINIITKKGDAEP-----GLHAEVTGGMGSGFAGSQDFD 191  
Qy 194 -SVAGV----SENGNVLARLDVYRTTGGAPDANGKRIAPDPAQTDKQDSKLSLNTVND 248  
Db 192 RNAAGAVTYNSENWD--ARLSIAGNRTGAFYDGSGLLIPDITQTSFAPNERIDLMSGIG 249  
Qy 249 WQLDDKQINIALTHYNDKQDTPADYDGNRLAVLFGKPSL-NAIKGLSLSEQPKTKTS 307  
Db 250 YQIDDDRRVEFSGGYFDSKQSDYGLYGPFFAAL--ADPSLFETRSGYESDFNPQTRRS 307  
Qy 308 TFINYHDDDLWGNTINTNAYVREKGRFYPFVA 341  
Db 308 MLNVYTTDNDVFGQOLLQSGSYRTERIKFHPFPA 341

RESULT 8  
A95420  
RhtA Rhizobactin receptor precursor [imported] - Sinorhizobium meliloti (strain 1021) ma  
C;Species: Sinorhizobium meliloti  
C;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
C;Accession: A95420  
R;Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bowe  
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.  
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001  
A;Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti  
A;Reference number: A95262; MUID:21396509; PMID:11481432  
A;Accession: A95420  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-746 <KUR>  
A;Cross-references: GB:AB006469; PIDN:AAK65923.1; PID:G14524437; GSPDB:GN00165  
A;Experimental source: strain 1021, megaplasmid pSymA  
R;Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, P.; Barloy-Hubler,  
Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A;Authors: Kahn, D.; Kahn, M.L.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
hebbaut, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A;Reference number: A96039; MUID:21368234; PMID:11474104  
A;Contents: annotation  
C;Genetics:  
A;Gene: rhtA  
A;Genome: plasmid

Query Match 22.1%; Score 453; DB 2; Length 746;

Best Local Similarity 32.6%; Pred. No. 3.9e-24;  
Matches 109; Conservative 67; Mismatches 132; Indels 26; Gaps 9;

Qy 18 AVTQOLYAQPNESTPVELEPVVITIDKSGMALANRITQMPHTTKVIYEEQIQEQATGSR 77  
Db 24 AVAQE---PANQSEAVTSLEIIVVTGGRS---AQQISEARTIYVVDSDQIQAEARSGK 76  
Qy 78 QLADVMAQLIPSLG-VSSGTTSNFGQTMHGQVQFLNGVPLTGSRDLSRQLNSINPNQV 136  
Db 77 TLQILGETIFSPDASDGARTSFQNLGRPLILVDGVSMSARSLSRQFDALDPENI 136  
Qy 137 ARIEVLGATSIYGGATGGLINIVTKSDLEEEQFETRIGVHGSKLSEGGIGYQVGQ--- 193  
Db 137 ERVEVLGATAIYGNATGGIINIITKKGDAEP-----GLHAEVTGGMGSGFAGSQDFD 191  
Qy 194 -SVAGV----SENGNVLARLDVYRTTGGAPDANGKRIAPDPAQTDKQDSKLSLNTVND 248  
Db 192 RNAAGAVTYNSENWD--ARLSIAGNRTGAFYDGSGLLIPDITQTSFAPNERIDLMSGIG 249  
Qy 249 WQLDDKQINIALTHYNDKQDTPADYDGNRLAVLFGKPSL-NAIKGLSLSEQPKTKTS 307  
Db 250 YQIDDDRRVEFSGGYFDSKQSDYGLYGPFFAAL--ADPSLFETRSGYESDFNPQTRRS 307  
Qy 308 TFINYHDDDLWGNTINTNAYVREKGRFYPFVA 341  
Db 308 MLNVYTTDNDVFGQOLLQSGSYRTERIKFHPFPA 341

RESULT 9  
AD1856  
hypothetical protein alr0397 [imported] - Nostoc sp. (strain PCC 7120)  
C;Species: Nostoc sp.  
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C;Accession: AD1856  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasanoto, S.; Watanabe, A.; Iriguchi,  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabei, S.; Tabei, S.;  
DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: AD1856  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-867 <KUR>  
A;Cross-references: GB:BA000019; PIDN:BAF72355.1; PID:gl7129742; GSPDB:GN00179  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: alr0397

Query Match 22.1%; Score 453; DB 2; Length 867;  
Best Local Similarity 33.5%; Pred. No. 4.8e-24;  
Matches 110; Conservative 63; Mismatches 143; Indels 12; Gaps 6;

Qy 14 LLSVAVTQOLYAQPNESTPVELEPVVITIDKSGMALANRITQMPHTTKVIYEEQIQEQ 73  
Db 165 ILATATEVAQDSAEVDEPIEL---VVTATRTPTIQN----VPRSTIVIDREQIAAQA 217  
Qy 74 TGSRLQADVMAQLIPSLGVSSGTTSNFGQTMHGQVQFLNGVPLTGSRDLSRQLNSINP 133  
Db 218 STSRNLIELKCTVPLGAPPAGASNGFLTGRNPQVLIDGVQSTTRNASRDLRTIDA 277  
Qy 134 NOVARIENVLSGATSIYGGATGGLINIVTKSDLEEE-QFETRIGVHGSKLSEGGIGYQVG 192  
Db 278 AAERIEVVRGPAIYDGTATGGTGGIINIITRPTBEKLTSTRTEVGSAAALGNLEGDSFSTN 337  
Qy 193 QSVAGVSENGNVLARLDVYRTTGGAPDANGKRIAPDPAQTDKQDSKLSLNTVNDVQL 251  
Db 338 LQHFISAKQGNFDTTFNFAVAKNGGFFDAQGDRIIPSDPAQGGFADASSINLFGKFGIDI 397  
Qy 252 DDKQINIALTHYNDKQDTPADYDGNRLAVLFGKPSLNAIKGLSLSEQPKTKTSTFNI 311  
Db 398 DANQRLQLTFRNRFDEKQDQTDIASD--PRVNTIPGRQKA-RALEGLSLDERPGNENTFINL 454



```
Qy 194 SVAGVSENGVNLRLDLDVYRTTGGAFDANGKRIAPEAQTDKQ-DSKSLSVNTNVDMQLD 252
Db 350 GISGQGGVDFIASFTRD--SEGTPTDAGDRI---PLFGDAEANSASINVLKGLGFLQG 404
Qy 253 DKQNINLALTHYNDKQDQTDYADPYGNRLAVLPGKPSNAIKGLSL-----SEQPKTKK 306
Db 405 SEQRLOITANYFENDQSDNV--DYD----LTVGQIPGQKARALDQPVFEINSTNPFNRG 458
Qy 307 STFNINYYHDDLWGMTNTINTNAYYRREK 333
Db 459 TVIQLDYTHNNILNSLOQAAYYRQTK 485

RESULT 13
Query Match 9.8%; Score 200; DB 2; Length 726;
Best Local Similarity 23.5%; Pred. No. 3.2e-06;
Matches 81; Conservative 65; Mismatches 134; Indels 64; Gaps 16;
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AG0837
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
T.; ConnerCon, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; PMID:11677608
A:Accession: AG0837
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-726 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD05886.1; PID:gl6503861; GSPDB:GN00176
C:Genetics:
C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolog

Query Match 9.8%; Score 200; DB 2; Length 726;
Best Local Similarity 23.5%; Pred. No. 3.2e-06;
Matches 81; Conservative 65; Mismatches 134; Indels 64; Gaps 16;
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AG0837
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
T.; ConnerCon, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; PMID:11677608
A:Accession: AG0837
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-726 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD05886.1; PID:gl6503861; GSPDB:GN00176
C:Genetics:
C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolog

Qy 7 FQWLSPLLSVAVTQQLVA--QPNESLPTVELEPVVITIDKSGMALANRITOMPHTTKVI 64
Db 8 FFWL-ITVSTGINSPLSAESTDNGETIVE-----STAEQVLKQPGVSI 55
Qy 65 YEEQIQEQATGSRQLADVMAQIPLSLGVSSTGTTNFGQTMHGRQVQF-----LLN 114
Db 56 TRDDIQKNP-VNDLADIRKM-PGVNLTNSAS--GTRGNRQIDIRGMGPENTLVLD 111
Qy 115 GVPLT-----GSRDISQLNSINPNQVARIIEVLG-ATSIYSGATGGLINIVTK 163
Db 112 GVPVTSRNSVRSYWRGERDTRGDTHWVPEMVERIEMIRGPAARYGSAAGGVVNIITK 171
Qy 164 SDLEEQFETRIGVHGSKLSSEGIGVQVQSVAGVSENGVNLRLDLDVY-RTTGGAFDAN 222
Db 172 RPTNDWHGSLSLTYNPPSSKEGDTFRGNFSLGSLGAGDTLTMRLYGNLNRLTADSWIN 231
Qy 223 G----KRIAPEAQTDKQDSKLSLVNTNVDMQLDKQNINLALTHYNDKQDQTDYADPYGN 278
Db 232 SSAGTKNAAGREGVTK-----DINSVFSWKMTQQILDPKAGY--SRQGNIVAGDTQN 283
Qy 279 RLAVLFGKPSNAIKGLSLSQPKTKSTFNINY--HHDDLWG 320
Db 284 -----SNSNAV-T-XSLAQSGRETNRNLYRQNYGLTHNGIWG 317

RESULT 14
AG2219
Hypochemical protein all3310 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AC2219
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
```

```
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG2219
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-698 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA075009.1; PID:gl7132405; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all3310

Query Match 9.2%; Score 187.5; DB 2; Length 698;
Best Local Similarity 25.0%; Pred. No. 2.3e-05;
Matches 91; Conservative 61; Mismatches 141; Indels 71; Gaps 21;
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AC2140
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2140
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-867 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA074373.1; PID:gl7131767; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2674
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homolog

Qy 6 YFOWLSPLLSVAVTQQL-----YAQPNESLPTVE--LEPVVITIDKSGMALANRIT 55
Db 39 YLSIELPTTSABLLTQSTPGDLNPEGEPEEQ-PEIETSSDDADITIE-----AIAEPET 93
Qy 56 QMPHT-TKVIYEEQIQEQATGSRQLADVMAQIPLSLGVSS---GTTNFGQTMHGRQVQ- 110
Db 94 LPASTPTVVIDQEEIQKQ--GATSVADVLRKM-PGFALNDVGHGADIHTGTYRGASINQ 150
Qy 111 --FLLGVPLTGSDDI-----SRQLNSINPNQVARIIEVLGATS-IYSGATGGLINIVT 162
Db 151 SVFLINGRRI--NNDVNTYHGATDLNLSIPVESIERVELSSGVTTSALYSSAFGQVNIIT 208
Qy 163 KSDLSEEQFETRIGVHGSKLSSEGIGVQVQSVAGVSENGVNLRLDLDV--YRTTGGAFD 220
Db 209 KKGVPQPKLTSLSEFSGSLNNLNQOFSY---SGAVGAATYNFSEFYPVDNRYRVPVGA-- 263
Qy 221 ANGKRIAPEAQTDKQ-----DSKSLSVNTNVDMQLDKQNINLALTHYNDKQDQTDY-- 272
Db 264 -----ANRDSQGLSNADTSTSTYFCNIGLGLDQRLNSLDITKLSRRGLVYFG 313
Qy 273 APDYGNRILAVLFGKPSNAIKGLSLSEQ-----PKTKSTFNINYYHDDLWGMTINT-N 326
Db 314 PFLQDRLL-----DHDGLNI--GLSWKTRLGNNGNSNLTTTFFGYNQNYFSTYGTPTVAGR 366
Qy 327 AYVR 330
Db 367 EFYR 370

RESULT 15
AC2140
ferrichrome-iron receptor [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AC2140
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2140
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-867 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA074373.1; PID:gl7131767; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all2674
C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homolog

Query Match 9.1%; Score 187; DB 2; Length 867;
Best Local Similarity 23.9%; Pred. No. 3.4e-05;
Matches 87; Conservative 52; Mismatches 131; Indels 94; Gaps 15;
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AC2219
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
```

```

Db      130 VSLPFTTSGTPPTTQPTTPTPTAETPTPETPTTPTPTAETPPPEEDEPTELIVGD 189
Qy      44 -----DKSGMALANRITOMPHTTKVIYEOIOEQATSGROLADYMAQLIPSLGVS 93
Db      190 QFEPPTYQVPMSSVSGTRDPTLDIVPQALQVIPQVETKQ--GTRSGELVKN----TSSAS 244
Qy      94 SGTSTN---FGOTMGRQVQPLINGVELTGSRDISRQLSINPNQVARIYVLGATSI- 148
Db      245 SGRTSSQAPALPTPIRGPESNNLLR---NGLRDQSLRQSSIAN--VERVEVLKGPASVYL 259
Qy      149 YSGATGGLINITYTKSDLBEEOFETRIGVHGSKLSESGIGYVGQ-----SVAGVSE 200
Db      300 FGGGDLGGVNVLTWKOPLNTPEY-----SIGYVGQGGGLHPRITDFGSLD 345
Qy      201 NGNVNVLARDVDYRTTGGAFDANGKRIAPPAQTDKQDSKSLSVNTNVMDQDDKQINL- 259
Db      346 KQGVAAKLINAAYQTA-----ESGKEDPENSESPFIAPVY--RLIGNENTNLT 389
Qy      260 ALTHYNDKQOTDVAPOYDGNRLAVLFGEKPSLNAIKGL---SLSE-QPKTKTSFNNIYHH 315
Db      390 ASIEYLKYRSEFETAPDLPASGTVISNPNGRVSRRETNLGEPSISESLVTRILGYQLDHRL 449
Qy      316 DDLW 319
Db      450 NDNW 453

```

RESULT 16

AC2134

ferrichrome-iron receptor [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp.

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002

C:Accession: AC2134

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Saemoto, S.; Matenabe, A.; Iriiguchi, N.; Nakazaki, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AC2134

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-872 <KUR>

A:Cross-References: GB:BA000019; PIDN:BA874325.1; PID:g17131719; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: alr2626

C:Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol

Query Match 8.9%; Score 182; DB 2; Length 872;  
Best Local Similarity 22.4%; Pred. No. 7.9e-05;  
Matches 78; Conservative 62; Mismatches 130; Indels 78; Gaps 16;

11 SLPLLSVAVLQQLAQPESLPTVELP-----VLTIDKSGMALAR--- 53

Db 163 AVETVITQTGPQIOTPFVEPQSTQTPBQPSAQGPAPAEILVGBQRVYPVNASTG 222

Qy 54 -----ITQPHHTTKVIYEQIQEQATGSRQLADYAAQL---IPSLGVSS-----GTTSN 99

Db 223 TRDTLTIRIDIPOTIOVPEQYIKQQRV--TRLRDALINTGGVAVQDGGFSGTSDQIGRGF 280

Qy 100 FGGTMRGRQVQFLANGVLTGSRDISRQLNSINPQVARIETVLSGATST-YSGGATGGGL 158

Db 281 FGGTGFGGSI--LVDFEK-DERGGIREFTAN-----VEEIEVLKAGSAVLYGGVQPGGVI 331

Qy 159 NIYTKSDLEEQEFTRIGVHSGKLSSEGIGYQVGQSVAG--VSENGVNLARLDVRYRTGG 217

Db 332 NIYTKQPLRDYVNAELSV--GSFST-----PRPSIDISGLPNSDKTLTLRLNSVYETSDG 365

Qy 218 AFDANGRIAPPAQTDKQDSKLSVNTNVMDQLDKONINITALHYNDKQDTD--YAP 274

Db 386 FRDFN-----QDVORFPIPLTKWEIGATVLTLLQFDVILINBERPRDGRGLAF 432

Oy 275 DYG---NRLAVLFGEKPSLNAIKGISLSEQPKTKSTFNINHYHDDLW 319  
 ||| : :: || : :: ||  
 Db 433 GEGIDTPLERPFGEIPDDVKYVERIGLS-----YRLHNFNNDW 471

RESULT 17  
AD2421  
ferriochrome iron receptor [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp.  
A:Note: Nostoc sp. Strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AD2421  
R:Kaneoko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Saito, M.; Watanabe, A.; Iriyuch  
N.; Kaneoko, T.; Shimo, S.; Sugimoto, M.; Takazawa, T.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium An  
A:Reference number: AB1807, MUID:21595285, PMID:11759640  
A:Accession: AD2421  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-858 <KUR>  
A:Cross-references: GB:BA000019, PDB:1BMB76623.1, PDB:1G17134062, GSPDB:GM00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
C:Gene: all4924

Query Match 8.9%; Score 181.5; DB 2; Length 858;  
Best Local Similarity 24.5%; Pred. No. 8,3e-05;  
Matches 91; Conservative 51; Mismatches 151; Indels 79; Gaps 17;

18 AVTQQLVAQNPESLPTV---ELAEFVITIDKSGMALANRITOMPHTKVIYEEOIQEAT 74  
Db 177 ALPEQPAQDPPEPIELVAVTEGQDQVRVPASVGRTRTDTPLADIQSIQVVRQVLDQOR- 235  
QY 75 GSRQLADVMAQOLI---PSLVSSGTTSNFGQTMGRGV---QFLINGVP---LTGSDDIS 125  
Db 236 -ATRGDLARVVGAVNPTRG--SGRADSF--TIRGFELFSGNVAVNNGIPDRLTETRD- 290  
QY 126 RQLNLSINNOVARIIVLSGATSI--YSGATGGLINITYTSDLSBEQFETRIGVSGKLS 184  
Db 231 -----YVEREVLKGPASVLYXGLNPGQYINITYKQLANPFNIEATVGNLYR 342  
QY 185 EGIGYQVQSAVGSSENGNVLARLDVDYRTTGAFDANKR---IAP----- 228  
Db 343 GGDIPS-----GPNDSKTLILYRLNLAIVQNSGYTIDFVGNRSFPIAVISAALDKNTT 397  
QY 229 -EPAOTDKQ-DKSLSVNTVNDMQL--DDKONIMLALTHYDKDDTDY-APDYGNRLAVL 283  
Db 398 FEGEVSQKTTISRIVVYVLAAGTVLPGPDGRRIRPRNRYVEPEDDTQIBTRLDYRLEHR 457  
QY 284 FGEKPSLNAIKGLSLSEOPKTKSTFNINHYHD-----DLMGNTINTNAYYRRE 332  
Db 458 FSENNSL-----RNDFRYTFEHMNDNNGAFLGLGDADNRNTARSTYSSES 502  
QY 333 KGRFYPVAPPS 344  
Db 503 DSNINLTITDIS 514

RESULT 18  
AC2079  
Ferrichrome-iron receptor [imported] - Nostoc sp. (strain PCC 7120)  
C|Species: Nostoc sp.  
A|Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C|Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C|Accession: AC2079  
R|Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kunitz, T.; Sasamoto, S.; Watanabe, A.; Iriuchih  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S  
DNA Res. 8, 205-213, 2001  
A|Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
C|Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AC2079  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-853 <KUR>  
A;Cross-references: GB:BA000019; PIDN:BA073884.1; PID:g17131276; GSPDB:GN00179  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: alr2185  
C;Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homology <TNC>

Query Match 8.8%; Score 180; DB 2; Length 853;  
Best Local Similarity 23.2%; Pred. No. 0.00011;  
Matches 82; Conservative 66; Mismatches 118; Indels 88; Gaps 19;

Qy 11 SLPLSVAVTQQLYAQPNESLFTVELEPV-VITIDKSGMALANRIT-----QMPH 59  
Dy 172 SLP-----PQPETQSEQSPAAE-EPIELVVTGKDGQVQVNTATVTRTNTPIIDIPO 224

Qy 60 TTKVYIEIQIQOATGSRQLADYMAQLIPSLGVSSGTTSNFQO--TMHGROVQVFLNGVP 117  
Dy 225 STQVVPQVLEDO-----QITRVDDALRNVPGLVGTNAFIGNQITIRG-----FTSNLP 275

Qy 118 LTGSRD---ISRLNSINPNQVARIEVLGATSI-YGSGATGGLINIVTKSDLEBEQPET 173  
Dy 276 IL--RDGRFRIYENFQETSNIERIEVLKGPASVOYQGLDPGVINLVTKPLSEPFVEI 333

Qy 174 RIGVHGSKLSSGIGYQVQGVQSVAG-VSENGVNLARLDVDYRTTGGAFDANGKRIAPAPAQ 232  
Dy 334 Q-----AQFGSYGL-IRPSFVSGPLTDDGKLLRLNATYQREGEFRDPN-----377

Qy 233 TDQKSKLSLVNTVDWQDDKQINLALTHYNDKQDTPDYADPNRLAVLFGKPSLNA 292  
Dy 378 ---TETERFFIAPSLTKISDRNTDFSLYLDSTRPFDTS-----LVAFGR-----421

Qy 293 IKGLSLSEOPKT-----TKS---TFNINVHDDLMGNTINTNAYYRRE 332  
Dy 422 ----SVADVPYSRVNPDPDFIDTKSFIAYNLEHFRSDNW--TLRNSPYLQQ 469

RESULT 19  
A40636  
ferric enterobactin receptor precursor Pfea PA2688 [imported] - Pseudomonas aeruginosa  
C;Species: Pseudomonas aeruginosa  
C;Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2000  
C;Accession: A40636; H83308  
R;Dean, C.R.; Poole, K.  
J. Bacteriol. 175, 317-324, 1993  
A;Title: Cloning and characterization of the ferric enterobactin receptor gene (pfea) of  
A;Reference number: A40636; MUID:93123148; PMID:8419284  
A;Contents: K407  
A;Accession: A40636  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-746 <DEA>  
A;Cross-references: GB:M98033; NID:g151434; PIDN:AAA25928.1; PID:g151435  
A;Note: sequence extracted from NCBI backbone (NCBI:122308, NCBI:122309)  
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bu  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A;Reference number: A82950; MUID:20437337; PMID:10984043  
A;Accession: H83308  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-746 <STO>  
A;Cross-references: GB:AE004597; GB:AE004091; NID:g9948760; PIDN:AAG06076.1; GSPDB:GN001  
A;Experimental source: strain PA01  
C;Genetics:  
A;Superfamily: ferricenterochelin receptor; tonB-dependent receptor amino-terminal homolo  
C;Keywords: membrane protein  
P;71-226/Domain: tonB-dependent receptor amino-terminal homology <TNN>

F;435-746/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

Query Match 8.8%; Score 179.5; DB 2; Length 746;  
Best Local Similarity 22.1%; Pred. No. 9.4e-05;  
Matches 96; Conservative 78; Mismatches 171; Indels 89; Gaps 17;

Qy 38 PVVITIDKSGMALAN-----RITQPHHTTKVYEEQIQEQATG-----SR 77  
Dy 7 PAVPFLLLSSCLLANAVHAAGQGVSVIEGQTVVATAQBEETKQAPGVSIITABDIAR 66

Qy 78 QLADVMAQLIPSL-GVSSGTTSNFQTMHGROVQ-----FLNGVPLT-----119  
Dy 67 PPSNDLSQIIRTPGVNLGTNSSSGQRGNRQIDIRGWPENTILLVDGKPVSSNSVRY 126

Qy 120 ---GSRDISROLNSINPNQVARIEVLG-ATSIYSGGATGGLINIVTKSDLEBEQFETRI 175  
Dy 127 GWRGERSRGTNNVPADQVERIEVIRGPAARVYNGAAGGVNIIITKQAGAEHTGNLSV 186

Qy 176 GVHGSKLSSGIGYQVQGVQSVAG-VSEN-----GNVLARLDVDYRTTGGAFDAN--GKRI 226  
Dy 187 YSNFPQHKACASERMSFGLNGLPTENLSYRVYGNII-AKTDSDDDWDINAGHESNRTGQA 245

Qy 227 APEPAQTDKQSKLSLVNTVDWQDDKQINLALTHYNDKQDTPDYADPNRLAVLFG 286  
Dy 246 GTLPA--GREGVRNKDIDGLSWRLTPEQTLFEAGF--SRQSNITGDTQN-----293

Qy 287 KPSLNAIKGLSLSEOPKTSTFNINVHDDLMGNTINTNAYYRREK-----333  
Dy 294 TSNNNYVKQMLGHETNRMRYRETSVTHRGWDVFGSSL---AYLQVEKTRNSRINEGLAGG 350

Qy 334 --GRFPYVAPFSTA--KALPILQSMNLP-----SATLDATKAPQARAYGLQSESAE 384  
Dy 351 TEGIFDPNNAGFYTATLRDLTAHGEVNLPLHLGYEQTLTLGSEWTEQKLDLDPSSNTQTE 410

Qy 385 VLGRVPLNKNPKRA 398  
Dy 411 EGGSIPLGAGKRS 424

RESULT 20  
AC2075  
outer membrane heme receptor alr2153 [imported] - Nostoc sp. (strain PCC 7120)  
C;Species: Nostoc sp.  
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C;Accession: AC2075  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: AC2075  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-802 <KUR>  
A;Cross-references: GB:BA000019; PIDN:BA073852.1; PID:g17131244; GSPDB:GN00179  
A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: alr2153

Query Match 8.7%; Score 178.5; DB 2; Length 802;  
Best Local Similarity 25.3%; Pred. No. 0.00012;  
Matches 94; Conservative 54; Mismatches 162; Indels 61; Gaps 17;

Qy 35 ELEPVVITIDKSGMALANRITQPHHTTKVYEEQIQEQATGSRQLADYMAQLIPSLGVSV- 93  
Dy 95 ENEEADIEITVGTTPRSLDESPTITVIDAEDIDNQLI---QNLDDLIRYEP--GVST 149

Qy 94 SGTTSNFG-----QTMHGROVQVFLNGVPLTGSRDII-SROL--NSINPNQVARIEVLG 144  
Dy 150 SGDARRYGFQDFNIRIGIDGNRVLLQVGVLPDSFDFGTQLGRNYIDTETLRRVEIRG 209

Qy 145 -ATSIYSGGATGGLINIVTK--SDLEEQFETRIGVHGSKLSSGIGYQVQGVQSVAGVSEN 201

Db 210 SASTLGVSDAIGVAVVFETKDPSPDYLNESGDD--GYFSNFVVDYDANRGIQNTTTLAIGRL 267  
 QY 202 GNVTLARDVAVYRTTGAFFDANGRIAPPAQTQKDSKLSVNTVNDWQDDKQNTNLAL 261  
 Db 268 GDVEGLL-ITTRDGEVPPQINSR-APNPQTTDAN-----SWTLKLVFNAGDDPQDLKJG 320  
 QY 262 THYNDDKQDTPADYDGNRLAVLFGKPSLNAIKGLSLSEQPKTKSTFNINHHDDLWGN 321  
 Db 321 EFINRTTDTNVLVSRG-----INFG-----VATDSLTATDEIKRKNRYNLSTEH-----N 364  
 QY 322 TINTNAVYRREKRFYFVAPFSIAKALPILOSKNLPASATLDAYTAQARAY----- 374  
 Db 365 NPDSNLFQVLRSGQIY-----YQEATSTESNELRRATAPITTGAVNRRRFRSSIIYQQ 417  
 QY 375 ----GVLOSSES 381  
 Db 418 NTFGGDIQLS 428

## RESULT 21

ORCFE  
 ferrichrome-iron receptor precursor - Escherichia coli (strain K-12)  
 C/Species: Escherichia coli  
 C/Date: 30-Jun-1988 #sequence\_revision 17-Oct-1997 #text\_change 01-Mar-2002  
 C/Accession: F64738; A25196; S45219; S06358  
 R/Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
 A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A/Title: The complete genome sequence of Escherichia coli K-12.  
 A/Reference number: A64720; MUID:97426617; PMID:9278503  
 A/Accession: F64738  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-747 <BLAT>  
 A/Cross-references: GB:AE00124; GB:U00096; NID:g1786339; PIDN:AACT3261.1; PID:g1786344;  
 A/Experimental source: strain K-12, substrain MGL55  
 J/Baculon, J.W.; Mason, P.; Cameron, D.R.; Carmel, G.; Jean, R.; Rode, H.N.  
 J. Bacteriol. 165, 181-192, 1986  
 A/Title: Protein functions of beta-galactosidase to the ferrichrome-iron receptor of Esche  
 A/Reference number: A25196; MUID:86085668; PMID:3079747  
 A/Accession: A25196  
 A/Molecule type: DNA  
 A/Residues: 1-608, 'RP', 611-747 <COU>  
 A/Cross-references: GB:D26562; NID:g473770; PIDN:BA05598.1; PID:g473809  
 A/Experimental source: strain K-12  
 R/Pujita, N.  
 submitted to the EMBL Data Library, January 1994  
 A/Reference number: S45219  
 A/Accession: S45219  
 A/Molecule type: DNA  
 A/Residues: 1-608, 'RP', 611-747 <FNU>  
 A/Cross-references: EMBL:D26562; NID:g473770; PIDN:BA05598.1; PID:g473809  
 A/Experimental source: strain K-12, substrain W3110  
 R/Burkhardt, R.; Braun, V.  
 Mol. Gen. Genet. 209, 49-55, 1987  
 A/Title: Nucleotide sequence of the fnuA and fnuH genes involved in iron (III) hydroxamate  
 A/Reference number: A32650; MUID:86038365; PMID:2823072  
 A/Accession: S06358  
 A/Molecule type: DNA  
 A/Residues: 723-747 <BUR>  
 A/Cross-references: EMBL:X05810  
 C/Genetics:  
 A/Gene: fnuA; tonA  
 A/Map position: 4 min  
 C/Function:  
 A/Description: located in the outer membrane, binds the ferrichrome-iron ligand; interac  
 system  
 C/Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homolog  
 C/Keyword: iron transport; membrane protein  
 F/1-33/Domain: signal sequence #status predicted <SIG>  
 F/34-747/Product: ferrichrome-iron receptor #status predicted <MAT>  
 F/98-235/Domain: tonB-dependent receptor amino-terminal homology <TMN>

F/468-747/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

Query Match 8.4%; Score 172; DB 1; Length 747;

Best Local Similarity 22.6%; Pred. No. 0.00032; Matches 91; Conservative 56; Mismatches 158; Indels 98; Gaps 19;

QY 5 HYFQWLSPLLSVAVTQOLYQAQ-----PNESLPVLEP-----VITIDKSGM 48  
 Db 11 HSLRKIAVVAVATVSGMSVYAAQAAVEPKEDTITVTAAPQESAMGPAATIAAQSGATG 70  
 QY 49 ALANRTQMPHTTKVYVEQIQEATGSRQLADYMAQLIPSGVSGT--TSNF----- 100  
 Db 71 KDTPIQKVPQISIVTAEM-----ALHQPKSVKALSYYTGVSVGTGASATYDHLI 125  
 QY 101 -GOTMGRQVQFLNGVPLTGS--RDISRLNSINNOVARIEVLSGATSI-YGSGATGG 156  
 Db 126 RGPAAEGQSONNYLNGLKIQGNFYND-----VIDPYLERAEMGPPSVLYGKSSPGQ 180  
 QY 157 LNIIVTK---SDLEBEQFETRIGVHSGKLSSEGIQYGVQSVAGVSENGVTLARLDVDY 212  
 Db 181 LNMVSKRPTEPLEKEVQF-----KAGTDSL-FQTGPDSLDLDGVYS-----Y 225  
 QY 213 RTTGAFFDANGRIAPPAQTQKDSKLSVNTVNDWQDDKQNTNLALTHNDKQDTPY 272  
 Db 226 RLITGLARSN-----AQKSGEQRVAIAAPFTMRPDDKNTFTF-LSYFQNEPETGY 276  
 QY 273 APYNGRLAVLFGKPSLNAIKGLSLSEQPKTKSTFNINHHDDLWGNNTNNAYRRE 332  
 Db 277 -----YGMLEKEGTVEPL---PNGKRLPTFN-----EGAKNNTYSRUB 312  
 QY 333 KGRFYPVAPFSIAKALPILOSKNLPASATLDAYTAQARAYG 375  
 Db 313 KQVYGFDFHFN--DIFTVRQNRF-----AKNTKSQSVYG 347

## RESULT 22

B85499  
 outer membrane receptor protein FnuA fnuA (imported) - Escherichia coli (strain O157:H7  
 C/Species: Escherichia coli  
 C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 17-May-2002  
 C/Accession: B85499  
 R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhe  
 iller, L.; Grobeck, E.J.; Davis, N.W.; Linn, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A/Reference number: AB5480; MUID:21074935; PMID:11206551  
 A/Accession: B85499  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-747 <STO>  
 A/Cross-references: GB:AE005174; NID:g12512871; PIDN:AA054454.1; GSPDB:GN00145; UNQGP:Z01  
 A/Experimental source: strain O157:H7, substrain EDL933  
 C/Genetics:  
 A/Gene: fnuA  
 C/Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homology  
 Query Match 8.4%; Score 171; DB 2; Length 747;  
 Best Local Similarity 22.3%; Pred. No. 0.00038; Matches 91; Conservative 57; Mismatches 152; Indels 108; Gaps 19;

Db 176 SSPGGLNNVSKRPTTEPLKEVQF-----KAGTSL-FQTGDFSDALDDGVYS- 224  
Qy 208 LDVDYRTTGGAFDANGKRIAPPAQTDKQDSKLSVNTNVDMLDQDKQINIALTHYNDK 267  
Db 225 ----YRLTGLARSAN-----AQQKSEEQRYAIAPAFTWRPDKNTFTF-LSYFQNE 271  
Qy 268 QDTDYADYGNRLAVLFGKPSLNAIKGLSLSEQPKTKSTNINYYHDDLLWGNTINTNA 327  
Db 272 PETGY-----YGLWPKRGTEVPL---PNGKRLPTDFN-----EGAKNNT 307  
Qy 328 YVREKGRFPYPVAPFSIAKALPILQSNLPSATLDAYTKAPQARAYG 375  
Db 308 YSRNEKMGVYSPDHEFN--DTFTVRQNLRP-----AENKTSQNSVYG 347  
RESULT 23  
B90648  
outer membrane receptor protein PhuA EC0154 [imported] - Escherichia coli (strain O157:  
C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 02-Aug-2002  
C:Accession: B90648  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
Saeawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: B90648  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-747 <H>  
A:Cross-references: GB:BA000007; PIDN:BAB33577.1; PID:gl3359610; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: EC0154  
C:Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homolog  
Query Match 8.4%; Score 171; DB 2; Length 747;  
Best Local Similarity 22.3%; Pred. No. 0.00038;  
Matches 91; Conservative 57; Mismatches 152; Indels 108; Gaps 19;  
Qy 5 HYFOWLSPLLSVAVTQOLYAQPNSLPTVELEPVVITI----- 43  
Db 11 HSLRKIAVVVATAVSGMSVIAO-----AAVELKEDTITVTAAPAPQESANGPAATIAAQ 65  
Qy 44 DKSGMALANRITQMPHTTKVIYEEIOEQATGSQLADVMAQLIPSLGVSSGT--TSNF- 100  
Db 66 SATGKTDTPIQKPSQISVVTAEEM----ALHQPKSVKEALSVTPGVSVGTGASNTY 120  
Qy 101 -----GQTMGRQVQVFLNGVPLTGS--RDISRQLNSINPNQVARIIVLSGATSI-YGS 151  
Db 121 DHLIIRGFAAEQSQNNVYGLKQGNFYND-----VIDPYMLERASIMRGPVSVLYGK 175  
Qy 152 GATGGLINIVTK-----SDLEEQFETRIGVHSGKLSSEGIGYQVQSGVAGSNGNVLAR 207  
Db 176 SSPGGLNNVSKRPTTEPLKEVQF-----KAGTSL-FQTGDFSDALDDGVYS- 224  
Qy 208 LDVDYRTTGGAFDANGKRIAPPAQTDKQDSKLSVNTNVDMLDQDKQINIALTHYNDK 267  
Db 225 ----YRLTGLARSAN-----AQQKSEEQRYAIAPAFTWRPDKNTFTF-LSYFQNE 271  
Qy 268 QDTDYADYGNRLAVLFGKPSLNAIKGLSLSEQPKTKSTNINYYHDDLLWGNTINTNA 327  
Db 272 PETGY-----YGLWPKRGTEVPL---PNGKRLPTDFN-----EGAKNNT 307  
Qy 328 YVREKGRFPYPVAPFSIAKALPILQSNLPSATLDAYTKAPQARAYG 375  
Db 308 YSRNEKMGVYSPDHEFN--DTFTVRQNLRP-----AENKTSQNSVYG 347  
RESULT 24  
B91346  
probable iron uptake protein Cj0755 [imported] - Campylobacter jejuni (strain NCTC 11168

C:Species: Campylobacter jejuni  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
C:Accession: D81346  
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel  
Nature 403, 665-668, 2000  
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
A:Reference number: A81250; MUID:20150912; PMID:10688204  
A:Accession: D81346  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-696 <PAR>  
A:Cross-references: GB:AL1139076; GB:AL111168; NID:g9686128; PIDN:CAB73021.1; PID:g9686820  
A:Experimental source: serotype O2, strain NCTC 11168  
C:Genetics:  
A:Gene: cfrA; Cj0755  
C:Superfamily: ferrienterochelin receptor; tonB-dependent receptor amino-terminal homolog  
Query Match 8.2%; Score 168.5; DB 2; Length 696;  
Best Local Similarity 25.0%; Pred. No. 0.00051;  
Matches 70; Conservative 57; Mismatches 104; Indels 49; Gaps 15;  
Qy 14 LLSVAVTQOLYAQPNSLPTVELEPVVITDKSGMALANRITQMPHTTKVIYEEIOEQ 73  
Db 14 LSSNAISQ-----NVELDSSIV----SASGFTQDIKEAPATINIVTKELQSKP 58  
Qy 74 TGSROLADVMAQLIP--SLGVSSGTTTNNFGQTMHG--RQVQFLNG-----VPLTGS 121  
Db 59 --YRDVAEATAD-IPGVDLVASKGKTGYNITMRTGTYTLVLIDGRRQGGVGPNGP 115  
Qy 122 ROISRQLNSINP--NOVARIIVLSGATS--YVSGATGGLINIVTKSLDEEOPTRIGVH 178  
Db 116 NEIS---NSFLPPISSIERIEVINGPMTLYGSEALGVVNIITKK--VSDKWETVSLD 170  
Qy 179 GSKLSSEGIGYQVQSGVAG-----VSENGNVLARDVDYRTTGA--FDANGKRIAPPAQ 232  
Db 171 ALLNENKDWNTYGTSTYSSGPLMNDKGLTLRPREYRQOSNVFEFTNGSQRVGQQAQ 230  
Qy 233 TDQDSKLSVNTNVDMLDQKQ----NINIALTHYNDKQ 268  
Db 231 SPTK-ANFNIGTRISYLANDYNTFIDFIDFSRNHYDNKQ 269  
RESULT 25  
D81257  
hemelin uptake system outer membrane receptor Cj1614 [imported] - Campylobacter jejuni (s  
C:Species: Campylobacter jejuni  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 03-Jun-2002  
C:Accession: D81257  
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrel  
Nature 403, 665-668, 2000  
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp  
A:Reference number: A81250; MUID:20150912; PMID:10688204  
A:Accession: D81257  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-709 <PAR>  
A:Cross-references: GB:AL1139079; GB:AL111168; NID:g9686971; PIDN:CAB73602.1; PID:g9686903  
A:Experimental source: serotype O2, strain NCTC 11168  
C:Genetics:  
A:Gene: chuA; Cj1614  
Query Match 8.2%; Score 168.5; DB 2; Length 709;  
Best Local Similarity 21.8%; Pred. No. 0.00052;  
Matches 74; Conservative 65; Mismatches 116; Indels 85; Gaps 15;  
Qy 60 TTKVIYEEIOEQATGSQLADVMAQLIPSL-----GVSSGTTTNSFG 101  
Db 26 SNKAINLQKVYVSTTGFEQDADSNLRNVISIEGKDLQNKGYVSLQALERISSISFVNF 85  
Qy 102 Q-----THGR-----QVQFLNGVP---LTGSRDISRQLNSINPNQVARIIVLSGATSI 148





A;Experimental source: strain PCC 7120  
C;Genetics:  
A;Gene: alr2175  
C;Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol

Query Match 8.1%; Score 165.5; DB 2; Length 863;  
Best Local Similarity 23.0%; Pred. No. 0.001;  
Matches 85; Conservative 59; Mismatches 154; Indels 71; Gaps 16;

Qy 10 LSLPLSVAVTQQLYAQPNEISLPTVELE---PVVITIDKSGMALNR-----ITQM 57  
Db 168 LIFTVASMATTAAQKPEEPQAPAAQDDPIELVVTGEADSVRVNAGAAATRTDPLRDI 227  
Qy 58 PHTTKYVEEQIQEATGSRQLADVMAQLIPSLGVSSGTTSNFGQ---TMHGRGV----- 109  
Db 228 PQSIQVPPQOVLKQOO-----VNTLNEALQNVSVGIQTASNYSQFASFTIRGFNSFDQG 281  
Qy 110 --QFLNGVPLTGSRDISRQLNSINPNQVARIEVLGATSI-YGSGATGGLINIVTKSDL 166  
Db 282 GNNFTRNGL---GYRFGSGGTFNS---IERIEVLGRGPGSVLFGSGNGPGGTINIVTKPL 335  
Qy 167 BEEQP--ETRIGVHSGKLSSEGIGYQVQGSVAG-VSENGNVLARLDVDYRTTGGAFDANG 223  
Db 336 SEPFYSVEAAGSYDF-----YRGALDLSGPLDDSKTALYRLNASYERKADNFVDEN- 386  
Qy 224 KRIAPEAQTQKQSKLSVNTNVDWQLDDKQINIALTHYNDKQDIDY---APDYGNR 280  
Db 387 -----DRENFVASTLSFALG--ENTTTLTLDQEVKNVQGYNGVPAVGTVL 431  
Qy 281 AVLFGKEPSLNAIKGLSSEQPKTKTSFNINYHHDDLWGNTINTNAYYRREKGRFPFV 340  
Db 432 PNLNGRIPNRSIGQADSTYSPEIVRVGYNLEHKFSEDW---LLRNARY---YSHFYNKT 485  
Qy 341 APFSIAKAL 349  
Db 486 RDTYFANAL 494

RESULT 29  
A12129  
ferrichrome-iron receptor [imported] - Nostoc sp. (strain PCC 7120)  
C;Species: Nostoc sp.  
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001  
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
A;Reference number: AB1807; MUID:21595285; PMID:11759840  
A;Accession: A12129  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-849 <KUR>  
A;Cross-references: GB:BA000019; PIDN:BA074291.1; PID:gl7131685; GSPDB:GN00179  
C;Experimental source: strain PCC 7120  
A;Gene: alr2592  
C;Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol

Query Match 8.0%; Score 163.5; DB 2; Length 849;  
Best Local Similarity 20.5%; Pred. No. 0.0015;  
Matches 80; Conservative 59; Mismatches 132; Indels 119; Gaps 16;

Qy 13 PLLSVAVTQQLYAQPNEISLPTVEPVVITIDKSGMALNRIT-----QMHTTKV 63  
Db 165 PPQTAQTEQAPAAQTEENQEPLEL---VVTGEQDIYVNPASTARTDPLRDI PQSIQV 221  
Qy 64 IYEQIQEATGSRQLADVMAQLIPSLGVSSGTTSNFGQTHGRGVQV-----L 112  
Db 222 IPQVLKDDQ-----VTRILDVAVNVSGVTPQKG---YGDATDFVTIRGFNSRNL 269  
Qy 113 LNVGVLTSGRDISRQLNSINPNQVARIEVLGATSI-YGSGATGGLINIVTKSDLSEEQF 171

Db 270 RGNFNVRAARGTSTFTTA--PSSVVERVELKGPASVLYGQLEPGGLVNIIVTKPLSDPY 327  
Qy 172 ETRIGVHSGKLSSEGIGYQVQGSVAG-VSENGNVLARLDVDYRTTGGAFD-ANGKRIAPE 229  
Db 328 AAETAGSYSP-----YSPSLDISGPLTTDKKLLYRLNASYQNGFSFIDPVNG----- 375  
Qy 230 PAQTDKQSKLSVNTNVDWQLDDKQINIALTH-----Y 264  
Db 376 -----ENLSIVPVVNYQISDRTALTTEYEHRSRNSQSFNDGLPIDPISFELPISRN 426  
Qy 265 NDKQDT-----DYAPDYGNRVLAVLFGKEPSLNAIKGLSLSQPKTKTKSTNIN----- 312  
Db 427 NNPDITLNTNSLNFLEHG-----FNDNIRLURTVFGASFSEGTTLAYRLFNYPNPTN 480  
Qy 313 -----YHDDLWGNTI-----NTNA 327  
Db 481 EVLAFRNPSPDYENDYSQWTDVIKFNPGA 510

RESULT 30  
S74447  
ferrichrome-iron receptor 1 - Synecchocystis sp. (strain PCC 6803)  
N;Alternate names: protein ell1403  
C;Species: Synecchocystis sp.  
A;Variety: PCC 6803  
C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
C;Accession: S74447  
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shingo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.  
DNA Res. 3, 109-136, 1996  
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecchocystis sp.  
A;Reference number: S74322; MUID:97061201; PMID:8905231  
A;Accession: S74447  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-863 <KAN>  
A;Cross-references: EMBL:D90899; GB:AB001339; NID:gl651650; PIDN:BAAL6599.1; PID:gl651651  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C;Genetics:  
A;Gene: fhua1  
A;Start codon: GTG  
C;Superfamily: ferrichrome-iron receptor 1; tonB-dependent receptor amino-terminal homol  
C;Keywords: iron transport  
F;240-375/Domain: tonB-dependent receptor amino-terminal homology <TNN>  
F;581-863/Domain: tonB-dependent receptor carboxyl-terminal homology <TNC>

Query Match 7.8%; Score 159.5; DB 2; Length 863;  
Best Local Similarity 23.3%; Pred. No. 0.003;  
Matches 84; Conservative 59; Mismatches 136; Indels 81; Gaps 17;

Qy 18 AVTQQLYAQPNEISLPTVEPVVITIDKSGMALNRITQMPHTTKYVEEQIQEATGSR 77  
Db 198 AAVGEFFV-PNTSVATGTDTPIMDT-----PFSQVQVSEEVIRSQ--AI 239  
Qy 78 QLADVMAQLIPSLGVSSGTTSNFGQTHGRGVQVFLNG-----VPLTGSRDISRQLN 129  
Db 240 TLEDVLT-----NVSSVT---FGTGTGRETIFGIRFGNGQFSDTVPIL--RDGFRLYG 288  
Qy 130 SI-----NPNQVARIEVLGATSI-YGSGATGGLINIVTKSLEBEQETRIGVHSGKLS 184  
Db 289 GFQGITVSHLQVQEVILKGPSSILYQIEPGGVINLNSKKPLNEPFAEVEV-----QLGN 343  
Qy 185 EGIGYQVQGSVAGVSENGNVLARLDVDYRTTGGAFDANG--KRIAPEAQTQKQSKLS 242  
Db 344 QCLVPRPFDISGGLNPNGLRYRLNGVYSNEASPRDNQPLERFAYAPI----- 392  
Qy 243 VNTNVDWQLDDKQINIALTHYNDKQDIDYA-PDYGNRVLAVLFGKEPSLNAIKGLSLEQ 301  
Db 393 ----VTVAITDDTDLAVEINDTNPADFLSGFGDGVAPV----PRSRVINDPSDIVN 444  
Qy 302 PKTKTSFNINYHHDDLWGNTINTNAYYRREKGRFPFVPAFPSTAKALPILQSNMLPSAT 361





## RESULT 36

E86022

outer membrane heme/hemoglobin receptor [imported] - Escherichia coli (strain O157:H7, #

C/Species: Escherichia coli

C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C/Accession: E86022

R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

Hiller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A/Reference number: A85480; MUID:21074935; PMID:11206551

A/Accession: E86022

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1660 &lt;STO&gt;

A/Cross-references: GB:AB005174; NID:G12518206; PIDN:AA658641.1; GSPDB:GN00145; UWGP:249

A/Experimental source: strain O157:H7, substrain EDL933

C/Genetics:

C/Superfamily: ferrichrome-iron receptor; tonB-dependent receptor amino-terminal homolog

Query Match 7.5%; Score 154.5; DB 2; Length 660;  
Best Local Similarity 25.4%; Pred. No. 0.0046;  
Matches 81; Conservative 50; Mismatches 119; Indels 69; Gaps 17;

QY 7 FQWTLPLSVAVTQQLYAPQNSLPVLEPVITIDKSGMALANITQMPHTTKYIYE 66

DB 6 FTSRLSLALAV-----SATLPFAFATETMTVTATNARS--FEAPMVSVYIDT 55

QY 67 EQLQEQATGSRQADVAQPLPSLGV--SSGTTNFGQTMHG---ROYQFLANGVPL--T 119

DB 56 SAPENQATG---ATDLRHVPGITIDGTGRTGQDVMMREYDRGLVLYVDGROGNDT 112

QY 120 GSRDISQLNS--INPNQAVIEVLGATG--IYSGATGGLI--NIYTKSLDEEQETTR 174

DB 113 G-----HNTGTFDPAIKVEIVRGPSALLYGSGALGVISYDVAKDILQEGSSG 166

QY 175 IGVHGS-KLSEGGIGYGVGSVAGVSENGVNLRLD-----VDYTTGAFANPKRIAPE 229

DB 167 FRVGTGTGTGSHSLG--LGSAFGRTEH-----LDGIWASSSDRGRLQSSNG--E 213

QY 230 PAQTDKQDSKSLSVNTVNDQLDKQINIALTHYNDKQTDVADPQNRRLAVLFGKSPS 289

DB 214 TAPRDEGINNNLAKGT---WQIDNAGSLSGVRYTND-----AAREPK 253

QY 290 LNAIKGLSLSEQPKTKST 308

DB 254 NQTVGASESSNPMVDRST 272

## RESULT 37

G90706

hypothetical protein Eca0623 [imported] - Escherichia coli (strain O157:H7, substrain R1

C/Species: Escherichia coli

C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 03-Aug-2001

C/Accession: G90706

R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gaeswara, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc

A/Reference number: A96229; MUID:21156231; PMID:11258796

A/Accession: G90706

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1746 &lt;HAU&gt;

A/Cross-references: GB:BA000007; PIDN:BA834046.1; PID:G13360081; GSPDB:GN00154

A/Experimental source: strain O157:H7, substrain R1MD 0509952

C/Genetics:

C/Superfamily: ferriterochelin receptor; tonB-dependent receptor amino-terminal homolo

Query Match 7.5%; Score 154.5; DB 2; Length 746;  
Best Local Similarity 21.9%; Pred. No. 0.0056;  
Matches 82; Conservative 61; Mismatches 129; Indels 103; Gaps 18;

QY 14 LLSVAVTQQLYAPQNSLPVLEPVITIDKSGMALANITQMPHTTKYIEEQIOEA 73

DB 11 LVNIGIVGAQAQEPDTPVSHDDTIVT-----AAEQNLQAPGVSTTADP-IRKNP 62

QY 74 TGSROLADVAQQLPSLGVSSGTTNFGQTMHGRQV-----FLANGVPLT----- 119

DB 63 V-ARDVSEII-RMPGVNLGNST--GQGNRRQIDIRMGEBENTLILIDGKPVSSRNS 118

QY 120 -----GSRDISQLNSINPNQAVIEVLG--ATSIYSGATGGLINITYTKSLDEEQE 172

DB 119 VRQGMGERTRDGTGVSVPPEMERIEVLKGPAAARVGNCAAGGVNIIIRKSGSGE---- 174

QY 173 TRIGVHGS-----KLSEGGIGYGVGSVAGV--SENGVNLRLDVDRYTTGAFDA 221

DB 175 -----WHQSMDAYENABEHKEGATKRTNFSLTGPLDDEFSFRLYGND--KTOADAMD1 227

QY 222 N-----GKRIAPPAQTDKQDSKSLSVNTVNDQLDKQINIALTHYNDKQTDV 273

DB 228 NQGHQARACTIVATYTPDPAGEGVYINK--DINGVRNDPAPLOSLEBAGY--SRQNLVA 283

QY 274 PD-----YGNRLAVLFGKSPSLNAIKGLSLSEQPKTKSTFNINYHDDLWG 320

DB 284 GDTQNTSDAYTRSKYGDENRLRYQNYSL-----TNWCG-----WD 320

QY 321 NTNTNAYTRREKGR 335

DB 321 NGVTSNMVQYEHTR 335

## RESULT 38

B83557

hypothetical protein fepA [imported] - Escherichia coli (strain O157:H7, substrain EDL9

C/Species: Escherichia coli

C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001

C/Accession: B83557

R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew

Hiller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A/Reference number: A85480; MUID:21074935; PMID:11206551

A/Accession: B83557

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1746 &lt;STO&gt;

A/Cross-references: GB:AB005174; NID:G12513473; PIDN:AA654918.1; GSPDB:GN00145; UWGP:207

A/Experimental source: strain O157:H7, substrain EDL933

C/Genetics:

C/Superfamily: ferriterochelin receptor; tonB-dependent receptor amino-terminal homolo

Query Match 7.5%; Score 154.5; DB 2; Length 746;  
Best Local Similarity 21.9%; Pred. No. 0.0056;  
Matches 82; Conservative 61; Mismatches 129; Indels 103; Gaps 18;

QY 14 LLSVAVTQQLYAPQNSLPVLEPVITIDKSGMALANITQMPHTTKYIEEQIOEA 73

DB 11 LVNIGIVGAQAQEPDTPVSHDDTIVT-----AAEQNLQAPGVSTTADP-IRKNP 62

QY 74 TGSROLADVAQQLPSLGVSSGTTNFGQTMHGRQV-----FLANGVPLT----- 119

DB 63 V-ARDVSEII-RMPGVNLGNST--GQGNRRQIDIRMGEBENTLILIDGKPVSSRNS 118

QY 120 -----GSRDISQLNSINPNQAVIEVLG--ATSIYSGATGGLINITYTKSLDEEQE 172

DB 119 VRQGMGERTRDGTGVSVPPEMERIEVLKGPAAARVGNCAAGGVNIIIRKSGSGE---- 174

QY 173 TRIGVHGS-----KLSEGGIGYGVGSVAGV--SENGVNLRLDVDRYTTGAFDA 221

DB 175 -----WHQSMDAYENABEHKEGATKRTNFSLTGPLDDEFSFRLYGND--KTOADAMD1 227

QY 222 N-----GKRIAPPAQTKQDSKSLSVNTVNDWQLDKONINLALTHYNDKQDTTDA 273  
DB 228 NOGHQSARAGTYATTPAGREGVINK--DINGVVRWDFAPLOSLELEAGY--SRQGNLYA 283  
QY 274 PD-----YGNRLAVLFGKPSLNAIKGLSLSEQPKTKSTFNINYYHDDLLWG 320  
DB 284 GDTQNTNSDAYTRSKYGDTRNLYRQNSL-----TWNGG-----WD 320  
QY 321 NTINTNAYRREKGR 335  
DB 321 NGVTTSNWWQVEHTR 335  
RESULT 39  
A83282  
probable TonB-dependent receptor PA2911 [imported] - Pseudomonas aeruginosa (strain PA01  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: A83282  
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
. J.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: A83282  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-718 <STO>  
A:Cross-references: GB:AE004717; GB:AE004091; NID:99948999; PIDN:AAG06299.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA2911  
Query Match 7.5%; Score 153; DB 2; Length 718;  
Best Local Similarity 27.2%; Pred. No. 0.0067;  
Matches 78; Conservative 43; Mismatches 102; Indels 64; Gaps 16;  
QY 10 LSLPLLSVAVTQQLYAQPNSLPTVELEPVVITIDKSGMALANRITOMPHTTKVIYEQI 69  
DB 35 LALPQLSVVRQ---DPAE-LDHIDLATPV-----SAGSLGLSALDTPASTSSISGEEV 85  
QY 70 QEQATGSQLADVMQAOLIPSLGVS--SGTTSNFGQTMHGR-----QVQFLNGVPL-TGS 121  
DB 86 RR-----RNNPSVQAAVTRSPGISPIGTGPGDGTLSARGFSGHASVMQLFDGTRLYTG- 139  
QY 122 RDISRLNSIN---PNOVARIEVLGATSI-YGSGATGGLINIVTKSDLEEE-QFETRI 175  
DB 140 -----MGTVNPSPDPMMVERIDVIRGPASVLYGEGATGAVINVPKPPFAGEIRNHURL 193  
QY 176 GVHGSKLSEGIYQVQGSV--AGVSENGNVLARDVDYRTTGGADFANGKRIAPEPAQT 233  
DB 194 G-YGS-----YDNRQLALDSGSLTDSLRYLNLNQQQSHG-----WI 230  
QY 234 DKQDSKSLSVNTVNDWQLDKONINLALTHYNDKQDTYAPDYGRL 280  
DB 231 DRGDSRNLGISAARWQASD-----DLAFTLAHDYGDQEPMDNFGTPL 273  
RESULT 40  
D87390  
TonB-dependent receptor [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C:Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: D87390  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwim, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A:Title: Complete Genome Sequence of Caulobacter crescentus.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: D87390

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-817 <STO>  
A:Cross-references: GB:AE005673; NID:gl3422452; PIDN:AAK23120.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC1136  
Query Match 7.5%; Score 153; DB 2; Length 817;  
Best Local Similarity 22.0%; Pred. No. 0.0081;  
Matches 92; Conservative 63; Mismatches 152; Indels 112; Gaps 21;  
QY 12 LPLLSVAVTQQLYAQPNSLPTVELEPVVITIDKSGMALANR-----ITQMPH 59  
DB 15 LAMPLAAHAQSVTKNDQ-DVLEDAVVITASRDGATKNNSSISVSALSADQLAQA 73  
QY 60 TTKVIYE--BOIQEQATGSQLADVMQAOLIPSLGVSSTTSNFGQTMHGRQVQFLNGV 117  
DB 74 STAEIFRGLPGVRSEATGGDNANIAVRGLP---VAGS-----GAKFLQLQEDGLP 121  
QY 118 LT-----GSRDISRLNSINPNOVARIE-VLSGATSIYSGNATGGLINIVTKSDLEEE 169  
DB 122 VMFEGDIAFGNADIFLRAD---QNVARIESVRGSGSSTFASNSPGGVINFIKSTG--- 173  
QY 170 QFETRIGVHGSKLS--SEGIGVQVQGSVAGVSENGNVLARDVDY-----RTTGAFD 220  
DB 174 -----EVEGEIATKGLGYD-----HTRLDPDYGAPLSDTLRPHVGGFY 213  
QY 221 ANGKRIAPEPAQTKQDSKSLSVNTVNDWQLDKQNTINLALTHYNDK-----QD 269  
DB 214 REGEGARKSGVTSEK--GGQIKANLTKDF---ENGFLNVLKVLNDRAVGLPMPTRVTG 268  
QY 270 TDYAPDYG-----NRLAVLFGKPSLNAIKGLSLSE-----QPKTKSTFNIN 314  
DB 269 SNGDPDYGAGYDIDGRDQLQSIYLAENPGLDGNRRISKVADGMHPDQLQGAERK- 327  
QY 315 HDDLWGNITNTNAYRREKGRFPYVAPF--SIKALPILQSMNLPSATLDAYTKAQA 371  
DB 328 --DIGWKIEDRFKAKTSGR---FVAPFPAEVLVSAQALATSIGGAGATL-RYANGPSA 380  
RESULT 41  
C98235  
probable hydroxamate-type ferrisiderophore receptor PA0470 [imported] - Agrobacterium tum  
C:Species: Agrobacterium tumefaciens  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
C:Accession: C98235  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, P.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A:Reference number: A97359; PMID:11743194  
A:Accession: C98235  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-710 <KUR>  
A:Cross-references: GB:AE007870; PIDN:AAK89405.1; PID:gl5159260; GSPDB:GN00170  
C:Genetics:  
A:Gene: AGR\_L\_1661  
A:Map position: linear chromosome  
Query Match 7.4%; Score 152; DB 2; Length 710;  
Best Local Similarity 24.8%; Pred. No. 0.0078;  
Matches 79; Conservative 45; Mismatches 131; Indels 64; Gaps 18;  
QY 32 PTVLEPVVITIDKSGMALANRITOMPHTTKVIYEEIQEQATGSQLADVMQAOLIPSLG 91  
DB 52 PKAPKGVKVAKTSASATKTSGLSVETQSVSVITKDQMDAQTV--RNLSEALNVPVGVVA 109  
QY 92 VSSGTTSNFG-----QTMHGRQVQFLNGVPLTGTSRDLSRQLNSINPNOV-----ARIEVL 144  
DB 110 QPSGADPRFADAPRIRGFGQNOLOP-LNGLRL-----MRTAGAPPVEVYGLERVEVIRG 161  
QY 145 ATSI-YGSGATGGLINIVTKSDLEEEQFETRIGVHGSKLSSEGIGYQ-----VGQSVAGVS 199

```

Db 162 PASVLYGGGNPGGLINLYSK-----RPTFE-RGEVGAQIGSPDY-YQSMFDIGGVASTD 215
Qy 200 ENGNYIARLDVDRYRTTGAFPDANGRIAPBPAQTDKODSKSLSVNTNVDWOLDKONIL 259
Db 216 -----SPAYRLTGLARKAH-----YQTNLNDQRFYIAPALTYQDEBETKLYV 258
Qy 260 ALTHYNDKODTDY----APDY---GNRLAVLF--GEKPSLNAIKGLSBEOPKTKSTFN 310
Db 259 LASYGHNDPSSSGLPPLALTYGRPNMMLDRSPYVD-FSPD-----TSNRKFTNIGYE 310
Qy 311 INYHDDLMGNTINTNAYY 329
Db 311 FEHRDETF--TFQONARY 327

```





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 25, 2002, 19:22:18 ; Search time 25 Seconds

(without alignments)  
663.621 Million cell updates/sec

Title: US-09-889-746-2

Perfect score: 2047  
Sequence: 1 MRSHYFQWLSPLLSVAVT.....SKAEVLGRVYNLKKPRALF 400

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	453	22.1	732	1	IUTA_ECOLI
2	453	22.1	746	1	RHTA_RHME
3	179.5	8.8	746	1	PFEA_PSEAE
4	172	8.4	747	1	FHUA_ECOLI
5	151.5	7.4	676	1	HMRU_YERPE
6	148.5	7.3	746	1	PEPA_ECOLI
7	144.5	7.1	652	1	IRGA_VIBCH
8	144.5	7.1	687	1	HEMR_YEREN
9	141	6.9	663	1	CIRA_ECOLI
10	134	6.5	815	1	FPVA_PSEAE
11	133	6.5	810	1	HPUB_NEIMC
12	130	6.4	723	1	Y262_HAEIN
13	129.5	6.3	700	1	YNCU_ECOLI
14	129	6.3	614	1	BTUB_SALTY
15	126.5	6.2	726	1	FATA_VIBAN
16	126	6.2	614	1	BTUB_ECOLI
17	124	6.1	735	1	FCT_BRNCH
18	123.5	6.0	735	1	HXC7_HAEIN
19	120.5	5.9	702	1	FOXA_SALTY
20	120	5.9	810	1	HPUB_NEIMA
21	118.5	5.8	512	1	FLAA_HELPE
22	118	5.8	908	1	TBL1_NEIMB
23	117.5	5.7	673	1	FYUA_YERPE
24	117	5.7	809	1	PUPB_PSEPU
25	116.5	5.7	944	1	LBPA_NEIMA
26	116.5	5.7	972	1	HGBA_HAEDU
27	115.5	5.6	673	1	FYUA_YEREN
28	114.5	5.6	760	1	YBIL_ECOLI
29	113	5.5	1306	1	FOXA_YEREN
30	113	5.5	1306	1	MSB2_YEAST
31	112.5	5.5	509	1	FLAA_HELPE
32	111.5	5.4	744	1	HXC1_HAEIN
33	110	5.4	729	1	PHUB_ECOLI

34	108	5.3	711	1	SP3_HUMAN
35	108	5.3	1169	1	C9DA_BACTP
36	107.5	5.3	972	1	HGBA_HAEDU
37	107.5	5.3	1729	1	TARP_HUMAN
38	106	5.2	694	1	DNLD_MYCLE
39	106	5.2	912	1	TBPI_HAEIN
40	106	5.2	106	1	LBPA_NEIMB
41	105.5	5.2	943	1	PUPA_PSEPU
42	105.5	5.2	826	1	BRUA_PSESP
43	104.5	5.1	340	1	HRPU_ERMCH
44	104.5	5.1	341	1	OMPL_PROPR
45	104.5	5.1	758	1	FCUA_YEREN

## ALIGNMENTS

RESULT 1  
ID IUTA\_ECOLI STANDARD; PRT; 732 AA.  
AC P14542;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Ferriic aerobactin receptor precursor (Cloacin receptor).  
GN IUTA.  
OS Escherichia coli.  
OC Plasmid IncFI ColV3-K30.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-R344;  
RA Krone W.J.A., Stegehuis F., Koningsstein G., van Doorn C.,  
RA Roosendaal B., de Graaf F.K., Oudega B.;  
RT "Characterization of the pColV-K30 encoded cloacin DF13/aerobactin  
RT outer membrane receptor protein of Escherichia coli; isolation and  
RT purification of the protein and analysis of its nucleotide sequence  
RT and primary structure";  
RL FEMS Microbiol. Lett. 26:153-161 (1985).  
RN [2]  
RP REVISIONS.  
RA Oudega B.;  
RL Submitted (FEB-1992) to the EMBL/Genbank/DDJ databases.  
CC -1- FUNCTION: RECEPTOR FOR CLOACIN DF13/AEROBACTIN.  
CC -1- SUBCELLULAR LOCATION: Outer membrane.  
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL: X05874; CAA29297.1; -  
DR EMBL: X05874; CAA29298.1; ALT\_INIT.  
DR PIR: S01042; S01042.  
DR InterPro: IPR000531; TonB boxC.  
DR Pfam: PF00593; TonB boxC.  
DR PROSITE: PS00430; TonB boxC.  
DR PROSITE: PS01156; TonB-DEPENDENT\_REC\_1.  
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Plasmid;  
KW Receptor.  
FT SIGNAL 1 25  
FT CHAIN 26 732 FERRIC AEROBACTIN RECEPTOR.  
FT SITE 31 38 TONB BOX.  
FT SITE 715 732 TONB C-TERMINAL BOX.  
SQ SEQUENCE 732 AA; 81014 MW; 0C23879C0B27AE2B CRC64;  
Query Match 22.1%; Score 453; DB 1; Length 732;

```

Best Local Similarity 30.8%; Pred. No. 2.3e-23;
Matches 133; Conservative 60; Mismatches 179; Indels 60; Gaps 9;

Qy 6 YFQW-----LSPLLSAVYVYQYLAQPNESLPTVELEPVVITIDKSGMALAVR-----ITQM 57
Db 7 YTLWALNPLLTMMAPAQAQ-----TDDTFVVSANRNRNRTVAEM 47

Qy 58 PHTTKVIYEEIOEQATGSRQLADWMAQLIPSLGVSSSTTSNFGCTMGROVQFLNCGVP 117
Db 48 AQTWVIEAELQIOGKELKDALAQLIFCLDVSSRSTYGNMVRGRPLVLVDGVR 107

Qy 118 LTGSRDISRLSNPNNOVARIEVLGATSIYGSATGGLNIVTKSDLEEQEPETRIGV 177
Db 108 LNSKTRDQRQSDSDPFNNHIEVIFGATSIYGGSTGGLNIVTKSQPPTWMEFEAGT 167

Qy 178 HGSKLSSGIGVQVQSGVAGVSENGNVLARDVDYRTTGGAPDANGKRIAPPAQTDKQD 237
Db 168 KSGFSSSKDHDEIAGAVSGGNE--HISGRLSVAYQKFGWFDGNGDATLLDNTQTGLQY 225

Qy 238 SKSLSVNTNVQWLDKQKINIALTHYNDKQDTPYAPDYGNRLAVLPCEKPSLNAIKGLS 297
Db 226 SDRLDIMGTLNIDESRQLQITQYKSGQDDYGLNLGKGFSAIRGTSTPFF-VSNGLN 284

Qy 298 LSEOPKTKSTFNINHYHDDLWGNNTINTNAVYRREKGRFYPP-----VAPFSIAKAL 349
Db 285 SDRIPGTGHLISLQYSDSAFLGQELVQVYRDESLRFPPTVNVANKQVATFSSQOD 344

Qy 350 PILQSMNLPSAT-----LDA-----YTKAPQARAYGVLOBESKAELVGR 388
Db 345 TDQYGMKLTLSKPMGWDGQITWGLDADHERFTSNQMPFLDAQASAGGLNNK-KIYTTGR 403

Qy 389 VPNLKPKRALP 400
Db 404 YPSVDITNLAAP 415

RESULT 2
RHTA RHIME STANDARD; PRT; 746 AA.
ID AC RHTA RHIME
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Rhizobium receptor precursor (TonB-dependent siderophore receptor
DE rhtA).
GN RHTA OR RA1265 OR SMA2414.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pSymA (megaplasmid 1).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RCR2011 / SU47;
RX MEDLINE=21172875; PubMed=11274118;
RA Lynch D., O'Brien J., Welch T., Clarke P., Cuiv P.O., Crosa J.H.,
RA O'Connell M.;
RT "Genetic organization of the region encoding regulation, biosynthesis,
RT and transport of rhizobactin 1021, a siderophore produced by
RT Sinorhizobium meliloti";
RN J. Bacteriol. 183:2576-2585(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubler F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federpiet N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid.";
RN Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).

```

```

CC -!- FUNCTION: RECEPTOR FOR THE SIDEROPHORE RHIZOBACTIN.
CC -!- SUBCELLULAR LOCATION: Outer membrane
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-aib.ch/announce/
CC or send an email to license@isb-aib.ch)
CC -----
CC EMBL; AF110737; AAD09419.1; -
CC EMBL; AE007312; AAK65923.1; -
CC IPR000531; TonB_boxC.1.
CC Pfam; PF00593; TonB_boxC.1.
CC PROSITE; PS00430; TONB DEPENDENT REC 1; FALSE NEG.
CC PROSITE; PS01156; TONB DEPENDENT REC 2; FALSE NEG.
CC Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW Plasmid; Complete proteome.
FT SIGNAL 1 26 POTENTIAL..
FT CHAIN 27 746 RHIZOBACTIN RECEPTOR.
FT SITE 40 47 TONB BOX.
FT SITE 729 746 TONB C-TERMINAL BOX.
SQ SEQUENCE 746 AA; 80633 MM; 16AB4A4A025D5B23 CRC64;

Query Match 22.1%; Score 453; DB 1; Length 746;
Best Local Similarity 32.6%; Pred. No. 2.3e-23;
Matches 109; Conservative 67; Mismatches 132; Indels 26; Gaps 9;

Qy 18 AVTQQLVAQNESLPTVELEPVVITIDKSGMALANRITQMPHTTKVIYEEIOEQATGSR 77
Db 24 AVAQ-----PANQSEAVTSLEEIVVTGGRS-----AQSEIARTIYVVDQIQAEASGK 76

Qy 78 QLADWMAQLIPSLG-VSSGTTTSNFGQTMHGRQVQFLNGVPLTGSRDISRQLNSINPQV 136
Db 77 TLQQLGETIPSPDASDGAFTSGQNLGRPPILLVDGVSMNSARSLSRQFADIDFNI 136

Qy 137 ARIEVLGATSIYGSATGGLNIVTKSDLEEQEPETRIGVHGSKLSEGIGYVQGV--- 193
Db 137 ERVEVLGATAIYGNATGGIINIITTKGKDAEP-----GLHAEVTGGMGSGFAGSQDDF 191

Qy 194 -SVAQV---SENGNVLARDVDYRTTGGAPDANGKRIAPPAQTDKQDSKLSVNTNVD 248
Db 192 RNAAGATVYSENWD--ARLSIAGNRIGAFYDGSGLTLLIDITOTSTAFNERIDLMGSI 249

Qy 249 WQLDDKQKINIALTHYNDKQDTPYAPDYGNRLAVLPCEKPSL-NAIKGLSLSEQPTTKS 307
Db 250 YQIDDDRRVEFGVQVFDKQSDYGLYGPFFAAL--ADPSLPETRSYGESDFNPQTRRS 307

Qy 308 TFINNYHDDLWGNNTINTNAVYRREKGRFYPPFA 341
Db 308 MLNVITYDNDVFGQQLLQGSYRTERIKFHPFPA 341

RESULT 3
PFEA PSEAE STANDARD; PRT; 746 AA.
ID AC PFEA PSEAE
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ferric enterobactin receptor precursor.
GN PFEA OR PA2688.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K407;
RX MEDLINE=93123148; PubMed=8419284;
RA Dean C.R., Poole K.;

```

RT "Cloning and characterization of the ferric enterobactin receptor  
 RT gene (flea) of *Pseudomonas aeruginosa*.";  
 RL J. Bacteriol. 175:317-324(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,  
 RA Reiter J., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
 RA Sailer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an  
 RT opportunistic pathogen.";  
 RT Nature 406:959-964(2000).  
 CC -1- FUNCTION: SPECIFIC RECEPTOR FOR THE SIDEROPHORE FERRIC  
 CC ENTEROBACTIN.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane.  
 CC -1- INDUCTION: BY IRON AND ENTEROBACTIN.  
 CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M80033; AAA25928.1; -;  
 DR EMBL; AE004697; AAG06076.1; -;  
 DR PIR; A40636; A40636.  
 DR HSSP; P05825; 1PEP.  
 DR InterPro; IPR000531; TonB\_boxC.  
 DR Pfam; PF00593; TonB\_boxC.1.  
 DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; FALSE\_NRG.  
 DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
 DR Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;  
 KW Complete proteome.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 746 FERRIC ENTEROBACTIN RECEPTOR.  
 FT SITE 39 44 TONB BOX.  
 FT SITE 729 746 TONB C-TERMINAL BOX.  
 SQ SEQUENCE 746 AA; 80967 MW; ADDAFB0CB2C351 CRC64;  
 Query Match 8.8%; Score 179.5; DB 1; Length 746;  
 Best Local Similarity 22.1%; Pred. No. 9.1e-05;  
 Matches 96; Conservative 78; Mismatches 171; Indels 89; Gaps 17;  
 QY 38 PAVITIDKSGMALAN-----RITQWPHTKVYEEQIQEQATG-----SR 77  
 DB 7 PAVPFLISSCLIANVAHAAGQDGSVIEIGEQVAVTAQEBKQAGVSIITAEADIAR 66  
 QY 78 QLAADVAAQLIPSLGVSSGTTNFGQTMGRQVQ-----FLINGVPLT----- 119  
 DB 67 PPNNDLSQIIRTPGVVLTJNSSSGQGNRRQIDIRMGEPENTLIIVDGKPVSSRNSVRY 126  
 QY 120 ---GSRDISQALNSINDQVARIIVLSG--ATSIYSGATGGLINITYKSLDLEBQFETRI 175  
 DB 127 GMRGERSRGDTWVPADQVERIEIVIRGPAARAGNAGGVNIIITKQAGAFTHGNLSV 186  
 QY 176 GVHGSKLSSSGIGVQVQSVAG--VSEN-----GNVIARLDVYRTTGAFAFDAN--GKRI 226  
 DB 187 YSNFPQKAKAGASERHSGFGLNGPITENLSTRVYGNITAKTSDSDWDINAGHSRFGKQA 245  
 QY 227 APBEAQTKDQSKSLSVNTVWQQLDDKQINILATHYNKQDTPADPYAGNRLAVLFG 286  
 DB 246 GTLPA--GRGAVNKKDIDGLSWRLTPEQULFEAGF--SRQGNITVGTQDN----- 293  
 QY 287 KPSLNAIKGLSLSEQPTTSTSTFNINYNHDDLNGNTINTNAYYRREK----- 333  
 DB 294 TNSNNVYKQMLGHETNMVREIYSVTHRGWDFGSSL---AYLQYETKRRSRINEGLAG 350

QY 334 --GRFPYFAPPSIA--KALPILQSNULP-----SATLDAYTKAPQARAYGVLSKESKAE 384  
 DB 351 TSGIFDPNNAGFYATLRLTALHGEVNLPLHIGYEQTLTIGSEWTEQKXDDPSNQTNE 410  
 QY 385 VLGKVPNLNKKPKRA 398  
 DB 411 EGSGISPLGAKKRS 424  
 RESULT 4  
 ID FHUA\_ECOLI STANDARD; PRT; 747 AA.  
 AC P06971; P75665; P71280;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Ferrichrome-iron receptor precursor (ferric hydroxamate uptake)  
 DE (ferric hydroxamate receptor).  
 GN FHUA OR TONA OR B0150.  
 OS *Escherichia coli*.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC *Escherichia*.  
 OX NCBI\_TaxID=562;  
 RX SEQUENCE FROM N.A., AND SEQUENCE OF 34-47.  
 RP STRAIN=K12;  
 RC MEDLINE=86085668; PubMed=3079747;  
 RX Coulton J.W., Mason P., Cameron D.R., Carmel G., Jean R., Jean H.N.;  
 RT "Protein fusions of beta-galactosidase to the ferrichrome-iron  
 RT receptor of *Escherichia coli* K-12.";  
 RL J. Bacteriol. 165:181-192(1986).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / W3110;  
 RX MEDLINE=94261430; PubMed=8202364;  
 RA Fujita N., Mori H., Yura T., Ishihama A.;  
 RT "Systematic sequencing of the *Escherichia coli* genome: analysis of  
 RT the 2.4-4.1 min (110,917-193,643 bp) region.";  
 RL Nucleic Acids Res. 22:1637-1639(1994).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=9742617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of *Escherichia coli* K-12.";  
 RL Science 277:1453-1474(1997).  
 [4]  
 RP SEQUENCE OF 482-647 FROM N.A.  
 RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,  
 RA Davis K., Federpiel N., Hyman R., Kaiman S., Komp C., Kurdi O.,  
 RA Lashkari D., Lew H., Lin D., Namach A., Oefner P., Robertes D.,  
 RA Davis R.W.;  
 RT Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.  
 RL [5]  
 RP SEQUENCE OF 723-747 FROM N.A.  
 RX MEDLINE=88038363; PubMed=2823072;  
 RA Burkhardt R., Braun V.;  
 RT "Nucleotide sequence of the fhuc and fhud genes involved in iron  
 RT (iii) hydroxamate transport: domains in fhuc homologous to  
 RT ATP-binding proteins.";  
 RL Mol. Gen. Genet. 209:49-55(1987).  
 RL [6]  
 RP REVIEW.  
 RX MEDLINE=94265918; PubMed=7515827;  
 RA Braun V., Killman H., Benz R.;  
 RT "Energy-coupled transport through the outer membrane of *Escherichia*  
 RT coli small deletions in the gating loop convert the fhua transport  
 RT protein into a diffusion channel.";  
 RL FEBS Lett. 346:59-64(1994).



```

Db 71 KDTPIQKVPQISVTAEM-----ALHOPKVEKALSTTPGVSGTRGASNTYDHLIT 125
Qy 101 -GQTMHGRQVQPLNGVPLTGS--RDISROLNSINPQVARIIEVLGATSI-YGSGATGG 156
Db 126 RGRAEBOGQSNNTYNGKLGKGNFYND-----VIDPYLERAIEKRPVSVLYKSSPGG 180
Qy 157 LNIIVTK---SDLEBOEFTFRIGVHSGKLSSEGIQVQSGVAGSENGVTLARLDVY 212
Db 181 LNMVSRKPTTEPLKEVOF-----KAGTDSL-FQTGFHFDSDLDGQVYS-----Y 225
Qy 213 RTTGAFADAGKRIAPPAQTDKODSKLSVNTVMDQDDKONINLALTHYNDKQDTDY 272
Db 226 RLVTGLASAN-----AQCKSSEORVYAIAPAFYWRPDKNTFT-LSYFQNPETGY 276
Qy 273 APDYGNLAVLFGKPSLNAIKGLSLSEOPPTTSTENINHYHDDLWGMTINTVAYYRR 332
Db 277 -----YGMLEPREGIVBL---PNGKRLPTDN-----EGAKNTTYSRNE 312
Qy 333 KGRFPYFVADPSPISAKALPILQSMNLPSPATLDAYTKAPQARAYG 375
Db 313 KMVGYSFDHEFN--DFTYRQNLRF-----AEKTKSQNSVYG 347

RESULT 5
HMUR_YERPE STANDARD; PRT; 676 AA.
ID HMUR_YERPE
AC 056989;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hemin receptor precursor.
GN HMUR OR YPO0283.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxId=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KIM6;
RX MEDLINE=96386041; PubMed=9026634;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Fellwell T., Hamlin N., Holroyd S., Jagers K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford A.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RA "Genome sequence of Yersinia pestis, the causative agent of plague,"
RA Nature 413:523-527(2001).
RL -1- FUNCTION: THIS PROTEIN IS INVOLVED IN THE INITIAL STEP OF IRON
UP TAKE BY BINDING HEMIN, AN IRON CHELATIN SIDEROPHORE THAT ALLOWS
THE BACTERIA TO EXTRACT IRON FROM THE ENVIRONMENT.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL; U60647; AAC64866.1; -.
DR EMBL; AJ414141; CAC89146.1; -.

```

```

DR InterPro; IPR00531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC.1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE NEG.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
KW Complete proteome.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 676 HEMIN RECEPTOR.
FT SITE 44 51 TONB_BOX.
FT SITE 659 676 TONB_C-TERMINAL_BOX.
SQ SEQUENCE 676 AA; 74230 MM; 84ED731CB914ACD3 CRC64;

Query Match 7.4%; Score 151.5; DB 1; Length 676;
Best local similarity 22.3%; Pred. No. 0.0063;
Matches 93; Conservative 64; Mismatches 187; Indels 73; Gaps 16;

Qy 4 SHYFQWLSPLSLVAATQQLYAOQNESLPV-----ELEPVITIDKSGMALANRTOM 57
Db 6 SDRFRWSSLS-LAIACQLPLATQADTTTQTSKHSSTDMVVT-----ATGNERSSP 58
Qy 58 PHTTKYIEEQIOEANGSROLADVMAQLPSIGVS-SGTTSNFGQTMGRQVQPLNGV 116
Db 59 EAPMWTYIEGNAPTQATATADMLRQ-VPGLVTVSGSKTNGQDVVMRGYKQGVLTIV 117
Qy 117 PLTGSRDISROLNS--INPQVARIIEVLSG-ATSIYSGATGGLI--NIVTKSLEEBOF 171
Db 118 DVGROGTDTGHNSTFLDPLALVARIIEVGRPALVLSGSLGSAIVAYETVADADMLQPGQ 177
Qy 172 ETRIIGHGSKLSSEGIQVQSGVAGSENGVTLARLDVYRTTGAFDAGKRIAPPA 231
Db 178 NSGVRVYSSAATGDH--SPGIGASAPFGRTDLDGI--LSFGRDIDIGNRQNGFN-----A 229
Qy 232 QTDKQKSKLSVNTVMDQDDKONINLALTHYNDKQDTDYADYGNRLAVLFGKPSLN 231
Db 230 PNDETISNVLAKGT--WQIDSIQSLSANLRYNN-----SAIEPKQPTIS 272
Qy 292 AIKGLSLSEOPPTTSTENINHY--HDDLWGMT-----INTNAVYRREKCR--- 335
Db 273 APSTTVMTNRSTIQDADQARKVIXKLDOEMLNATQVYSEVINARPGSSEEGREOT 332
Qy 336 -----FYFPVAPSPISAKALPILQSMNLPSPATLDAYTKAPQARAYGVLQSE 380
Db 333 TEGVKLENTRFLFIESPASHLTYGTETVYQEOETPGGATESFPOAKIRFSSGLQBE 389

RESULT 6
FEPA_ECOLI STANDARD; PRT; 746 AA.
ID FEPA_ECOLI
AC P05825; P75722; P76821; P77093;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ferrienterobactin receptor precursor (Enterobactin outer-membrane
DE receptor).
GN FEPA OR FEP OR FEUB OR B0584.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxId=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=86278160; PubMed=3015941;
RA Lundtigan M.D., Kadner R.J.;
RA "Nucleotide sequence of the gene for the ferrienterochelin receptor
RA FePa in Escherichia coli. Homology among outer membrane receptors
RA that interact with TonB,"
RL J. Biol. Chem. 261:10797-10801(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

```



RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,  
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,  
RA Fraser C.M.;  
RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*  
RT cholerae";  
RL Nature 406:477-483(2000).  
RN (3)  
RP SEQUENCE OF 1-151 FROM N.A.  
RC STRAIN=classical Ogawa 395 / ATCC 39541 / Serotype O1;  
RX MEDLINE=91072235; PubMed=2174861;  
RA Goldberg M.B., Boyko S.A., Calderwood S.B.;  
RT "Transcriptional regulation by iron of a *Vibrio cholerae* virulence  
RT gene and homology of the gene to the *Escherichia coli* fur system.";  
RL J. Bacteriol. 173:6863-6870(1990).  
CC -1- FUNCTION: INVOLVED IN THE INITIAL STEP OF IRON UPTAKE BY BINDING  
CC PERIC VIBRIOBLACTIN, AN IRON CHELATIN SIDEROPHORE THAT ALLOWS  
CC V. CHOLERA TO EXTRACT IRON FROM THE ENVIRONMENT.  
CC -1- SUBCELLULAR LOCATION: Outer membrane.  
CC -1- MISCELLANEOUS: REGULATION OF THE IRGA EXPRESSION IS NEGATIVELY  
CC REGULATED AT THE TRANSCRIPTIONAL LEVEL BY IRON.  
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U72152; AAC4766.1; -;  
CC EMBL: AE004134; AAF93648.1; -;  
CC PIR: A37834; A37834.  
CC PIR: S25265; S25265.  
CC HSSP: P05825; 1FEP.  
CC TIGR: VC0475; -;  
CC InterPro: IPR000531; TonB\_boxC.  
CC DR Pfam: PF00593; TonB\_boxC.1.  
CC DR PROSITE: PS00430; TONB\_DEPENDENT\_REC\_1; 1.  
CC DR PROSITE: PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
CC KM Virulence; Outer membrane; Iron transport; Transport; TonB box;  
CC Signal; Receptor; Complete proteome.  
CC FT SIGNAL 1 25 POTENTIAL.  
CC FT CHAIN 26 652 IRON-REGULATED OUTER MEMBRANE VIRULENCE  
CC PROTEIN.  
CC FT SITE 33 40 TONB BOX.  
CC FT SITE 635 652 TONB C-TERMINAL BOX.  
CC FT CONFLICT 294 294 D -> G (IN REF. 1).  
CC FT CONFLICT 448 448 K -> Q (IN REF. 1).  
CC FT CONFLICT 502 502 A -> T (IN REF. 1).  
CC FT CONFLICT 502 502 A -> T (IN REF. 1).  
CC SQ SEQUENCE 652 AA; 71669 MW; A95P82FEC072EC93 CRC64;  
  
Query Match 7.1%; Score 144.5; DB 1; Length 652;  
Best Local Similarity 22.9%; Pred. No. 0.018;  
Matches 83; Conservative 47; Mismatches 114; Indels 119; Gaps 18;  
  
QY 37 EPVYITIDKSGMALANITQMPHTTKYIERQIQEQATGSGQLADWAQ--ISLSGVSS 94  
DB 33 EIMVYI---AAGYAYQYQNAFASISVYSRDL-----SKYRDVTDALSKSVGVTVTG 83  
  
QY 95 GTTSNFGQT---MHGRVQFLNGVPLTGSRDISRQ--NSINP-----NQVAR 138  
DB 84 G-----GDTTDISIRGMSNVTLL--ILVDGKRQTSRQRPSPDGGIISGWLPLQAIER 136  
  
QY 139 IEVLSGATG--IYSGGATGGLINIVTKSDLEBEQPTRIIGHGSKLSSBGIGVQGVAG 197  
DB 137 IEVIRGSMSTLYGSDAIGVINIITRKD--QQQWSGVN-----QLSTV 177  
  
QY 198 VSENGVTLARLDVYRTTGAFAFDANGKRIAPAPOTKQD-----SKSL-SVNTNVDWQ 250  
DB 178 VQENRAGSDGDSANFPYTPGISDALSLQVYGGQTTQRDEDEIEHGYGDSLSLSTSKANYQ 237  
  
QY 251 LD-----DKQN-----INLATHYNDKQDPTDY 272

DB 238 INPDHQLQLEAGVSAADRENWGSQAQSSGCRGTGNTDQYRRNHVAVASHQDMDQVQ 297  
QY 273 APYGNRLAVLPFKKPSLNAIKGLSLSEQKTKTKST-----FNINHHDDLWGN 321  
DB 298 SDTY-----LQYBE-----NTNKSREMSIDNTVFKSLVADIGHMUSFGVEGHSLEDK 348  
  
QY 322 TIN 324  
DB 349 TSN 351  
  
RESULT 8  
HEMR YEREN STANDARD; PRT; 687 AA.  
AC P31459;  
DT 01-UTL-1993 (Rel. 26, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DE 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Hemlin receptor precursor.  
GN HEMR.  
OS Yersinia enterocolitica.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Yersinia.  
OX NCBI TaxID=630;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 51872 / WA-C / Serotype O:8;  
RX MEDLINE=93049186; PubMed=1425573;  
RA Stojiljkovic I., Hanke K.;  
RT "Hemin uptake system of *Yersinia enterocolitica*: similarities with  
RT other TonB-dependent systems in Gram-negative bacteria";  
RL EMBO J. 11:4359-4367(1992).  
RN [2]  
RP REVISIONS.  
RC STRAIN=ATCC 51872 / WA-C / Serotype O:8;  
RA Stojiljkovic I.;  
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.  
CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN THE INITIAL STEP OF IRON  
CC UPTAKE BY BINDING HEMIN, AN IRON CHELATIN SIDEROPHORE THAT ALLOWS  
CC THE BACTERIA TO EXTRACT IRON FROM THE ENVIRONMENT.  
CC -1- SUBCELLULAR LOCATION: Outer membrane.  
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X68147; CAA48250.1; -;  
CC DR PIR: S28042; S28042.  
CC DR InterPro: IPR000531; TonB\_boxC.  
CC DR Pfam: PF00593; TonB\_boxC.1.  
CC DR PROSITE: PS00430; TONB\_DEPENDENT\_REC\_1; 1.  
CC DR PROSITE: PS01156; TONB\_DEPENDENT\_REC\_2; FALSE\_NG.  
CC KM Outer membrane; Iron transport; Transport; TonB box; Signal;  
CC Receptor.  
CC FT SIGNAL 1 28 POTENTIAL.  
CC FT CHAIN 29 687 HEMIN RECEPTOR.  
CC FT SITE 44 51 TONB BOX.  
CC FT SITE 670 687 TONB C-TERMINAL BOX.  
CC SQ SEQUENCE 687 AA; 75226 MW; 69092EF975DC08A CRC64;  
  
Query Match 7.1%; Score 144.5; DB 1; Length 687;  
Best Local Similarity 22.3%; Pred. No. 0.019;  
Matches 96; Conservative 61; Mismatches 174; Indels 99; Gaps 18;  
  
QY 4 SHYFQWISLPLSLVAATVQYQNPNSLPTVE-----LEPVYITI--DKSGMALANRI 54  
DB 6 SDRFRWSPIS-LAIACITSLAVQAAVTSSTQTNSSKRIADTMVTVATGNRSSEFAPMV 64

QY 55 TOMPHITTKVIYEEQIOEQATGSRQLADYMAQLIPSLGV-SGTTSNFQOTMHRGQVPLL 113  
 Db 65 TVVEADTPT-----SETATS-----ATMLRNIPGLVTGSRVNGQDVTLRGYKQGV 114  
 QY 114 NGVPLTGSRDLSROLNS--INPNQVAREVLVSGATS-IYSGCATGGLINIVTKSLDLEE 170  
 Db 115 TLVDGIROGTDTGHLNSTFLDPALVKRVEIVRGSALLYGSGALGGVIS----- 163  
 QY 171 PETRIGVHGSXLSSEGGYQVQGSVA-----GVSENGNVNVLARLDVYRTTGGAFDAN 222  
 Db 164 YET-VDAADLLLPQNSGCVRYVYSAATGDHSGFLGASAFRT---DDVDGILSFGTRDIG 219  
 QY 223 GKRIAPEAQTQKDSKSLSVNTNV-----DWOLDKQINIALTHYNDKQDTPDYADPYN 278  
 Db 220 NIR-----QSDGFNAPDETISNVLAKGTWRIDQIQLSANLRYNN----- 261  
 QY 279 RLAVLFGCKPSLNAIKGLSLSEQPKTKSTFNINVHDDLAGNTNTNA--YYRR----- 331  
 Db 262 --SALEPKNPOTSAASTNLMTDRSTIORDAQLKYNIKPLDQEWLNATAQVYSEVEINA 319  
 QY 332 -----EKGR-----FYFPVAPFSIAKALPILQSMNLSATLDAYTKAPQ 370  
 Db 320 RPQGTPEGRKQTTGGKLENKTRLTFTDSFASHLLTYGTEAYKQQTSPGATESPQADI 379  
 QY 371 ARAYGVQLOSE 380  
 Db 380 RFGSGWLQDE 389

## RESULT 9

CIRA\_ECOLI STANDARD; PRT; 663 AA.  
 AC P17315;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Colicin I receptor precursor.  
 GN CIRA OR CIR OR FEUA OR B2155.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89123100; PubMed=2644220;  
 RA Nau C.D., Konisky J.;  
 RT "Evolutionary relationship between the TonB-dependent outer membrane  
 transport proteins: nucleotide and amino acid sequences of the  
 Escherichia coli colicin I receptor gene.";  
 RL J. Bacteriol. 171:1041-1047(1989).  
 RN [2]  
 RP REVISIONS.  
 RA Nau C.D., Konisky J.;  
 RL J. Bacteriol. 171:4530-4530(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / BHB2600;  
 RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,  
 Church G.M.;  
 RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [5]  
 RP SEQUENCE OF 1-59 FROM N.A., AND SEQUENCE OF 26-46.

RX MEDLINE=88058737; PubMed=3316180;  
 RA Griggs D.W., Tharp B.B., Konisky J.;  
 RT "Cloning and promoter identification of the iron-regulated cir gene  
 of Escherichia coli.";  
 RL J. Bacteriol. 169:5343-5352(1987).  
 RN [6]  
 RP SEQUENCE OF 1-125 FROM N.A.  
 RC STRAIN=BPR2;  
 RX MEDLINE=92250419; PubMed=1315732;  
 RA Steffes C., Ellis J., Wu J., Rosen B.P.;  
 RT "The lypS gene encodes the lysine-specific permease";  
 RL J. Bacteriol. 174:3242-3249(1992).  
 RN [7]  
 RP SEQUENCE OF 1-21 FROM N.A.  
 RX MEDLINE=90264362; PubMed=2160948;  
 RA Griggs D.W., Kafka K., Nau C.D., Konisky J.;  
 RT "Activation of expression of the Escherichia coli cir gene by an  
 iron-independent regulatory mechanism involving cyclic AMP-cyclic  
 AMP receptor protein complex.";  
 RL J. Bacteriol. 172:3529-3533(1990).  
 CC -1- FUNCTION: NOT YET KNOWN. POSTULATED TO PARTICIPATE IN IRON  
 TRANSPORT. OUTER MEMBRANE RECEPTOR FOR COLICINS IA AND IB.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane.  
 CC -1- INDUCTION: BY IRON AND BY A CYCLIC AMP/CYCLIC AMP RECEPTOR  
 PROTEIN COMPLEX.  
 CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; J04229; AAA75183.1; -;  
 DR EMBL; U00007; AAA60531.1; -;  
 DR EMBL; AB000304; AAC75216.1; -;  
 DR EMBL; M19295; AAA23581.1; -;  
 DR EMBL; M89774; AAA17054.1; -;  
 DR PIR; A32056; QREIC.  
 DR PIR; S24561; S24561.  
 DR HSP; P05925; IPEP.  
 DR EC02DBASE; B068.1; 6TH EDITION.  
 DR EcoGene; EG10155; cirA.  
 DR InterPro; IPR000531; TonB\_boxC.  
 DR Pfam; PF00593; TonB\_boxC; 1.  
 DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; 1.  
 DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
 KW Outer membrane; iron transport; Transport; TonB box; Signal; Receptor;  
 KW Complete proteome.  
 FT SIGNAL 1 25  
 FT CHAIN 26 663 COLICIN I RECEPTOR.  
 FT SITE 31 38 TONB BOX.  
 FT SITE 646 663 TONB C-TERMINAL BOX.  
 FT CONFLICT 97 97 S -> D (IN REF. 6).  
 FT CONFLICT 528 528 I -> N (IN REF. 1).  
 SQ SEQUENCE 663 AA; 73895 MW; 2C68A45D4B5EE414 CRC64;  
 Query Match 6.9%; Score 141; DB 1; Length 663;  
 Best Local Similarity 21.2%; Pred. No. 0.032; Indels 90; Gaps 12;  
 Matches 65; Conservative 53; Mismatches 99;  
 QY 40 VITIDKSG-----MALANRITQMPHTTKVIYEEQIOEQATGSRQLADYMAQLIPSLGV 92  
 Db 23 VLAVDDGDETTVVTVASSVEQNKLKDPASISVITQEDLQRPV--QNLKDVLEK-VPGVQL 79  
 QY 93 SSGTTSNFGQTMHGRQVQLNGVPLTGSRIISR-----QLNSINPNQVARIIVLS 143  
 Db 80 TNEGDNKRGVSGIRGLDSSYTL--ILVDGKRVNSRNAVFRHNDFDLNIWIPVDSIERIEYVR 137  
 QY 144 G-ATSIYGSATGGLINIVTKSLDLEEEQFETRIGVHGSXLSSEGGYQVQGSVA----- 197



```

Db 138 GPMSSLYSDALGAVNIITK-----XIGKMSGTVTD 171
Oy 198 --VSENGVILARDVDYRTTG-----CAFANGKRIAPBPACTKQSKSL----- 241
Db 172 TTIOHRRGDGTYNQGFISGPLIDGVLGKMGKYGSLAREKDDPONSPTTDTGETPRIEG 231
Oy 242 --SVNTAVDWQDDQONINIALTHYNDKODPD-----YAP-----DYGRL 280
Db 232 PSSRGKGVFPATPNQNHDPFTAGYGFDDKDRBDSLSLDKRLERONYSVSHGRMDYGTSE 291
Oy 281 AVLFGKX 287
Db 292 LKYGEX 298

RESULT 10
FPVA_PSEAE STANDARD; PRT; 815 AA.
ID PVA_PSEAE
AC P486132; OS1339;
DT 01-FEB-1996 (Rel. 33, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ferripyoverdine receptor precursor.
GN FPVA OR PA2398.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 44-64.
RC STRAIN=CD10;
RX MEDLINE=93328663; PubMed=8335619;
RA Cloning K., Neshat S., Krebes K., Heinrichs D.E.;
RT "Cloning and nucleotide sequence analysis of the ferripyoverdine
  receptor gene fpyA of Pseudomonas aeruginosa.";
RN J. Bacteriol. 175:4597-4604 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PAO;
RA Laminot I.L.;
RN Submitted (MAR-1994) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Bryn A.L., Mizoguchi S.D., Warren P.,
  Hickey M.J., Britman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
  Garber R.L., Goltzy L., Tolentino E., Westbrock-Neckman S., Yuan Y.,
  Brody L.L., Coulter S.N., Folger K.R., Kas A., Labdig K., Lam R.M.,
  Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
  Reiter J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
  opportunistic pathogen.";
RN Nature 406:959-964 (2000).
RN [4]
RP FUNCTION: RECEPTOR FOR THE SIDEROPHORE, FERRIPYOVERDINE.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- INDUCTION: BY THE SIDEROPHORE, PYOVERDINE, AND UNDER IRON
  STARVATION CONDITIONS.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-16 IS THE INITIATOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration -
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL, L10210; AAA25819.1; -
DR EMBL, U07359; AAB60199.1; -
DR EMBL, AE004666; AAG05786.1; -
DR PIR, A40601; A40601.

```

DR InterPro:IPR00531; TonB boxC.  
DR Pfam: PF00593; TonB\_boxC\_1.  
DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; FALSE\_NEG.  
DR PROSITE; PS01155; TONB\_DEPENDENT\_REC\_2; 1.  
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;  
KW Complete proteome.  
FT SIGNAL 1 43 POTENTIAL.  
FT CHAIN 44 815 PERIPEROXIDINE RECEPTOR.  
FT SITE 798 815 TONB C-TERMINAL BOX.  
FT CONFLICT 716 815 Y -> F (IN REF. 1).  
FT CONFLICT 745 746 MISSING (IN REF. 1).  
FT CONFLICT 750 750 S -> R (IN REF. 1).  
SO SEQUENCE 815 AA; 91167 MM; A11A84A5A290F35 CRC64;

Query Match 6.5%; Score 134; DB 1; Length 815;  
Best Local Similarity 21.7%; Pred. No. 0.13;  
Matches 80; Conservative 63; Mismatches 166; Indels 60; Gaps 16;

QY 33 TVELEPVVITIDKSG-----MALANRIT-----QMPHTTKYIYEQIQEATNG 75  
DB 124 SVDLGATMTSNQIGTITEDSGSYTPGTIATPTRLVLTPTRETPSITVVTRONMDD--FG 181  
QY 76 SROLADVAALQILPSIGVSGSTTNSGCTMGHGRGV-QFILANGVPLVSGRDISRQL-NSINP 133  
DB 182 LAMIDDVAKR-TFGITTSAYITDRNNRYARAFSPINNFYDGIPTI-ANNVGSAGNTISD 239  
QY 134 NQV-ARIEVLGATSIY-GSGATGGLINIVTKSLDEEOPETRIQVHSGSKLSECIQYV 191  
DB 240 MAIDYREVLGATGTLGAGSLGATINILIKKPTHEPKHVEIG-----AGSWDYRS 293  
QY 192 GQSVAG-VSENGVTLARDVYRTTGGCAFDPANGKRIAPRPATQDKQDSKLSVNTNVDMQ 250  
DB 294 ELVDGSPLTESGNRGRAVVAAYODKHSFMD-----HYERKTSVYVYGI--- 335  
QY 251 LDDKQINILATLHYNDKQDTPADPYGRLAVLFEKXPSLNAI-----KGLSLSEQPKTT 305  
DB 336 LEFDLNPPTMLTGVADYDNDPKSGSGMSGSPFLPFGSQNRDVSFSFNNGAKWSMEQYT 395  
QY 306 KSTF-NINVHHDDLMGNTI---MTNAYVREKGRFPYFPAFPISTAKALPILOSNNLPSA 360  
DB 396 RTVFANLEHNPNANGVGVQDLHKINGHAPLGAIMGWPAPDNSAKILVAQKYTGETSNN 455  
QY 361 TLDAYTKAP 369  
DB 456 SLDIYLTGP 464

RESULT 11  
HPUB\_NEIMC HPUB\_NEIMC STANDARD; PRT; 810 AA.  
AC P66949;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemoglobin-haptoglobin utilization protein B precursor.  
GN HPUB.  
OS Neisseria meningitidis (serogroup C).  
CC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI\_Taxid=135720;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DNM2 / Serogroup C / Serotype 2a;  
RX MEDLINE=97206152; PubMed=9157745;  
RA Lewis L.A., Gray E., Wang Y.-P., Roe B.A., Dyer D.W.;  
RT "Molecular characterization of hpuaB, the haemoglobin-haptoglobin-  
utilization operon of Neisseria meningitidis".  
RL Mol. Microbiol. 23:737-749(1997).  
CC -1- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE  
HEMOGLOBIN/HAPTOGLOBIN COMPLEX AND IS REQUIRED FOR HEME UPTAKE.  
CC -1- SUBCELLULAR LOCATION: Outer membrane (potential).  
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U73112; AAC44893.2; -.  
DR InterPro; IPR000531; TonB\_box.  
DR Pfam; PF00593; TonB\_box; 1.  
DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; FALSE\_NEG.  
DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
KW Outer membrane; transport; TonB Box; Signal; Receptor.  
FT SIGNAL 1 22 POTENTIAL.  
FT CHAIN 23 810 HEMOGLOBIN-HAPTOGLOBIN UTILIZATION  
FT SITE 793 810 TONB C-TERMINAL BOX.  
FT SEQUENCE 810 AA; 90626 MW; 0D08DD79DA9CB817 CRC64;

Query Match 6.5%; Score 133; DB 1; Length 810;  
Best Local Similarity 22.5%; Pred. No. 0.15;  
Matches 100; Conservative 62; Mismatches 164; Indels 118; Gaps 27;

QY 13 PLLSVAVTQOLYAQNESLPTVELEPVITIDKSGMALANRITOM-PHTTKVLYEEIOQE 71  
DB 7 PVLAVAAIAQAF-----PAPAADPA-----PQSAQTL-NEITVTGTHKTQKLGEEKIR 54  
QY 72 QATGSRQLADV--MAQLIPSLG-VSSGTTSNFGQTMHG---RQVQFLNGVPLTGSR--- 122  
DB 55 KTLDKLLVNDHDLVRYDPCISVVEGGRAGSNGFTIRGVDKRVAINVDGLQAESRSSE 114  
QY 123 -----DISRLQNSINPNOVARIEVLGATSI-YGSGATGGLINIVTK--SDLEEE 169  
DB 115 AFOELFGAYGNFNAARNTSEPFSEVITTKGADSLKSGSGALGGAVNYQTKSADYVSE 174  
QY 170 QETRIGVHG-----SKLSEG-----IGV--QVGQSVAGSENGVNLARD 209  
DB 175 DKPYHLGKGGVSGKNSQKFSSTIAGRLFLGDLALVYTRFRGKETKWRSTGNEVEIKND 234  
QY 210 ---VDVRTTGG-----AFDANG-KRIAPEPAQ--TDKQDSKLSLVNTN-----VDWQLDDKQ 255  
DB 235 GVVNPPTDGGSKYLYTYATGVARSQPDQEWVNSKTLFLKGYNFNDQNRIGWIFEDSR 294  
QY 256 -----NINLALHYNDKQTDY-----APDYGNRLAVLFGKEKSLNAIKGL 296  
DB 295 TDRFTNELSNLWTGTTTSAATGDYRHRQDVSRYRRRSRGSVEYKNEI-----EHGPMDSLKL 349  
QY 297 SLSEQPKTKSTFNI--NYHDDDLGNTINTNAY--YRREKGRFPFVAPFSAIKALPIL 352  
DB 350 YDKQIDMNTWTWDIPKNY---DLRG--INSEVYHSFRHIRQNTAQWTDADF----- 395  
QY 353 QSMNLPSATLDAYTKAPQARAYGV 376  
DB 396 -----EKQLD-PSKAVAAQVGL 412

RESULT 12  
Y262\_HAEIN  
ID\_Y262\_HAEIN STANDARD; PRT; 723 AA.  
AC P44600;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Probable tonB-dependent receptor HI0262 precursor.  
GN HI0262.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID:727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-Rd / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Karlavage A.R., Bult C.J., Tomb J.-P., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Udelman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Weierback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
RT Rd.";  
RL Science 269:496-512(1995).  
RN [2]  
RP IDENTIFICATION BY MASS SPECTROMETRY.  
RX MEDLINE=20137488; PubMed=10675023;  
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,  
RA Gray C., Fountoulakis M.;  
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";  
RL Electrophoresis 21:411-429(2000).  
CC 1- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT.  
CC 1- SUBCELLULAR LOCATION: Outer membrane (Potential).  
CC 1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; U32712; AAC21927.1; -.  
DR TIGR; HI0262; -.  
DR InterPro; IPR000531; TonB\_box.  
DR Pfam; PF00593; TonB\_box; 1.  
DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; FALSE\_NEG.  
DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
KW Outer membrane; Receptor; Signal; TonB box; Complete proteome.  
FT SIGNAL 1 221 POTENTIAL.  
FT CHAIN 22 723 PROBABLE TONB-DEPENDENT RECEPTOR HI0262.  
FT SITE 706 723 TONB C-TERMINAL BOX.  
SQ SEQUENCE 723 AA; 80775 MW; 91EB3AB0FEEA2984 CRC64;

Query Match 6.4%; Score 130; DB 1; Length 723;

Best Local Similarity 24.1%; Pred. No. 0.2;  
Matches 84; Conservative 51; Mismatches 137; Indels 76; Gaps 19;

QY 7 FOWLSLPLLSVAVTQOLYAQNESLPTVELEP--VVITIDKSGMALANRITOMPHTTKVI 64  
DB 3 FSKLSLAITTTLTANALAQ-----SVELDSINVIAIRDPSS-----RFAYTPE--KQS 48  
QY 65 YEEQIQOQATGSRQLADVMAQLIPSLGVSSGTTSS-----NFGQTMHGQVQLNGVPLT 119  
DB 49 KDSLLSKQAT---SVADALEP-IPNVVGRGSRSTIAQPNIRGLSDNRVWQ-VIDGV--R 101  
QY 120 GSRDISRLQNSINP-NOVARIEVLGSG-ATSIYSGATGGLINIVTKSDLEESQFETRIGV 177  
DB 102 QNFDLAHRGSYFLPMSLIQIEIVIKGPPSSSLGSGALGGVAVMRTPNALDLLKNDKFGV 161  
QY 178 HGSKLSSEGIGYQVGQSVAGVSENGVNLARDVDVYRTTGGAFDA--NGKRIAPEAQTDK 235  
DB 162 ---KIRQ---GYQ-----TANNLSEKDVSVFAANDKEDVLISGFYNNADNLRTCK 205  
QY 236 QDS-----KSLSVNTNVDMQLDDKQNTINLALHYNDKQTDYAPDYGNRLAVLFGKEPS 289  
DB 206 GNKLANTAYKQFGLAKFGWQINDANRVELSHRETRFKQTA-----PS 248  
QY 290 LNAIKGLSLSEQPKTKSTFTNTNYHDDDLGNTINTNAYYRREKGRFY 337  
DB 249 NNEVENELTNEQ--ITDQIKKXPHGQKDDLLPPTTPQSP---SERSEFY 291

RESULT 13

YNCB\_ECOLI  
ID YNCB\_ECOLI STANDARD; PRT; 700 AA.  
AC P76115;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DN Probable tonB-dependent receptor yncD precursor.  
OS YNCB OR B1451.  
OC Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_Taxid=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Colado-Vides J., Glaeser J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).  
CC -1- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT.  
CC -1- SUBCELLULAR LOCATION: Outer membrane (potential).  
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: AE000242; AAC74533.1; -  
CC EcoGene; EG13774; yncD.  
DR InterPro: IPR000531; TonB\_boxC.  
DR Pfam: PF00593; TonB\_boxC; 1.  
DR PROSITE: PS00430; TONB\_DEPENDENT\_REC\_1; FALSE NEG.  
DR PROSITE: PS01156; TONB\_DEPENDENT\_REC\_2; FALSE NEG.  
KM Hypothetical protein; Outer membrane; Receptor; Signal; TonB box;  
KW Complete proteome.  
FT CHAIN 1 23 POTENTIAL.  
FT SITE 24 700 PROBABLE TONB-DEPENDENT RECEPTOR YNCB.  
FT SITE 680 700 TONB C-TERMINAL BOX.  
SQ SEQUENCE 700 AA; 77260 MW; 7B3B96C6ABA48F64 CRC64;  
  
Query Match 6.3%; Score 129.5; DB 1; Length 700;  
Best Local Similarity 22.3%; Pred. No. 0.21;  
Matches 57; Conservative 49; Mismatches 103; Indels 47; Gaps 11;

BTUB\_SALTY  
ID BTUB\_SALTY STANDARD; PRT; 614 AA.  
AC P37409;  
DT 01-OCT-1994 (Rel. 30, Created)  
DT 15-JUN-1994 (Rel. 30, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DN Vitamin B12 receptor precursor.  
OS BTUB OR STM4130.  
OC Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_Taxid=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=L72;  
RX MEDLINE=93079446; PubMed=1448622;  
RA Wei B.Y., Bradbeer C., Kadner R.J.;  
RT "Conserved structural and regulatory regions in the Salmonella  
RT typhimurium btub gene for the outer membrane vitamin B12 transport  
RT protein.";  
RL Res. Microbiol. 143:459-466(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=L72 / SGGC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Molyneay E.,  
RA Ryan E., Sun H., Florea L., Miller W., Steneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
RT L72.";  
RL Nature 413:852-856(2001).  
CC -1- FUNCTION: COBLAMIN RECEPTOR PROTEIN. ALSO ACTS AS A RECEPTOR FOR  
CC BACTERIOPHAGE BP23 AND IS NECESSARY FOR THE UPTAKE OF B COLICINS.  
CC -1- SUBCELLULAR LOCATION: Outer membrane.  
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M89481; AAA27031.1; -  
CC EMBL: AE008893; AAL22968.1; -  
CC SLYGene; SG10030; btub.  
DR InterPro: IPR000531; TonB\_boxC.  
DR Pfam: PF00593; TonB\_boxC; 1.  
DR PROSITE: PS00430; TONB\_DEPENDENT\_REC\_1; 1.  
DR PROSITE: PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
KM Outer membrane; Cobalt transport; Transport; TonB box; Signal;  
KW Receptor; Complete proteome.  
FT CHAIN 1 20 BY SIMILARITY.  
FT SITE 21 614 VITAMIN B12 RECEPTOR.  
FT SITE 26 33 TONB BOX.  
FT SITE 597 614 TONB C-TERMINAL BOX.  
FT CONFLICT 81 81 G -> R (IN REF. 1).  
FT CONFLICT 134 134 R -> P (IN REF. 1).  
FT CONFLICT 183 183 N -> I (IN REF. 1).  
FT CONFLICT 280 280 R -> S (IN REF. 1).  
SQ SEQUENCE 614 AA; 66525 MW; 9F51F601A615F662 CRC64;  
  
Query Match 6.3%; Score 129; DB 1; Length 614;  
Best Local Similarity 24.7%; Pred. No. 0.19;  
Matches 67; Conservative 42; Mismatches 108; Indels 54; Gaps 10;

QY 76 SRQLADVMAQL-----IPSLGVSSGTTSNFGQTMHGRQVQFLLNGVPL-----TGSRDLSR 126  
 Db 59 STSVNDVLRRLPGVDIAQSGGAGQNSIFIRGTNSHVLVLDIGVRLNLAVGSGSADLSQ 118  
 QY 127 QLSNPNQVARIIEVLGATS-IVSGATGGLINIVTKSDLEEEQFETRIGVHSGKLSSE 185  
 Db 119 FVSL-----VRIEIVIRPRAIYSDAIGGVNIIITRDNGPTELTAGWSNSYQNYDI 174  
 QY 186 GIGVQVGSVAGSVENGVLARLDVYRTTGGAFDANGRIAPPAQTDKQDSKLSVNT 245  
 Db 175 STQQLGE-----NTRATLIGDYETKGF-DVVAKGGTGMAQDPDRDGLSKTLYG 224  
 QY 246 NVDWQLDKQINLALHYNKQDID--YAP 274  
 Db 225 ALEHTFSDRWSGFVRGYDNRDIDYAYSP 255

RESULT 15  
 FATA\_VIBAN  
 ID FATA\_VIBAN STANDARD; PRT; 726 AA.  
 AC P11461; P19830;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1989 (Rel. 12, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ferric anguibactin receptor precursor (OM2).  
 GN FATA.  
 OS *Vibrio anguillarum* (*Listonella anguillarum*).  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Listonella.  
 OX NCBI\_TaxID=55601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=775;  
 RX MEDLINE=88139336; PubMed=2830268;  
 RA Actis L.A., Tolmasek M.E., Farrell D.H., Croas J.H.;  
 RT "Genetic and molecular characterization of essential components of  
 the *Vibrio anguillarum* plasmid-mediated iron-transport system.";  
 RL J. Biol. Chem. 263:2853-2860(1988).  
 RN [2]  
 RP SEQUENCE OF 637-726 FROM N.A.  
 RC STRAIN=775;  
 RX MEDLINE=90185247; PubMed=2311935;  
 RA Farrell D.H., Mikesell P., Actis L.A., Croas J.H.;  
 RT "A regulatory gene, *angr*, of the iron uptake system of *Vibrio*  
*anguillarum*: similarity with phage p22 *cro* and regulation by iron.";  
 RL Gene 86:45-51(1990).  
 CC -1- FUNCTION: ANGUIBACTIN RECEPTOR, ESSENTIAL COMPONENT OF THE IRON  
 UPTAKE SYSTEM OF *V. ANGUILLARUM*.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane.  
 CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; J03529; AAA91581.1; -;  
 CC EMBL; M34504; AAA79859.1; -;  
 CC PIR; B29228; B29928.  
 CC PIR; PQ0051; PQ0051.  
 CC InterPro; IPR000531; TonB\_boxC.  
 CC Pfam; PF00593; TonB\_boxC\_1.  
 CC PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; 1.  
 CC PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
 CC Signal; Receptor; Transmembrane; Outer membrane; Iron transport;  
 KW Transport; TonB box; Plasmid.  
 FT SIGNAL 1 35 POTENTIAL.  
 FT CHAIN 36 726 FERRIC ANGUIBACTIN RECEPTOR.  
 FT SITE 49 56 TONB BOX.  
 FT SITE 709 726 TONB C-TERMINAL BOX.

SQ SEQUENCE 726 AA; 78894 MW; 88867A219395B154 CRC64;  
 Query Match 6.2%; Score 126.5; DB 1; Length 726;  
 Best Local Similarity 21.4%; Pred. NO. 0.35;  
 Matches 82; Conservative 63; Mismatches 165; Indels 73; Gaps 17;  
 QY 10 LSLPLLSVAVTQ-----QLYAQNESLPTVLEPVTITDKSCMALANRITQMPHT 60  
 Db 31 ISASALPISITHAEQEADSEITVYQANEAYAAKISKAKA-----SSIGMLGDKDFLDPTFN 86  
 QY 61 TKVYEEQIQEQTGSRQLADVMAQLIPSLGVSSGTTSNFGQTMHGRQVQFLLNGVPLTG 120  
 Db 87 AIGYTDKHIQDQ--HAQDISDNVISASDPVS-FTSGETGLNKESPKIRGFSDDIGDVMPNG 143  
 QY 121 SRDISQLNSINPNQVARIIEVLGATSIGY-----SGATGGLINIVTK-----S 164  
 Db 144 LYGIAPYVRS-SPEMYQRIDVLKGPASLLMGMPNGSVGGGSLNLTVKRAQEAIPITSTGT 202  
 QY 165 DLEEEQFETRIGVHSGKLSSEGIGYQVQSVAGVSENGNV-----LARDVDYRT 214  
 Db 203 YMSDSQFGGHIDI-GRRF---GENEFGVRFNGVFRDGDASVDGQSRKAQLASLSDWRN 258  
 QY 215 TGGAFDAN-----GKRFA-----PEPAQTDKQDSKLSVNTNVDWLDKQON 256  
 Db 259 DIALIEDLYFSTERVDPNRLGSLASGVDPVSPSSDTLLSPSWAYNDSEDKMMIRAE 318  
 QY 257 INL--ALHYN--DKQDTPADYAPYGNRLAVLFGKPSLNAIKGLSLEQPKTKTKSTFNIN 312  
 Db 319 LDLNSVTAYGAVCASRTDPDSNVQVRKII-DDSGTLEVSIG-SVKLESKRTSGEVGIR 376  
 QY 313 YHHD--DLWGNNTINTNAYRREK 333  
 Db 377 SSFDTGPIEHLVNLNSTYFREDK 399

RESULT 16  
 BTUB\_ECOLI  
 ID BTUB\_ECOLI STANDARD; PRT; 614 AA.  
 AC P06129;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Vitamin B12 receptor precursor.  
 GN BTUB OR BFE OR CER OR DCRC OR B3966.  
 OS *Escherichia coli*.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OX *Escherichia*.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=85130824; PubMed=3882670;  
 RA Heller K., Kadner R.J.;  
 RT "Nucleotide sequence of the gene for the vitamin B12 receptor protein  
 in the outer membrane of *Escherichia coli*.";  
 RL J. Bacteriol. 161:904-908(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=94089392; PubMed=8265357;  
 RA Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,  
 RA Daniels D.L.;  
 RT "Analysis of the *Escherichia coli* genome. IV. DNA sequence of the  
 region from 89.2 to 92.8 minutes.";  
 RL Nucleic Acids Res. 21:5408-5417(1993).  
 RN [3]  
 RP SEQUENCE OF 1-5 FROM N.A.  
 RX MEDLINE=9115413; PubMed=1999392;  
 RA Gustafsson C., Lindstrom P.H., Hagervall T.G., Esberg B.,  
 RA Bjoerk G.R.;  
 RT "The *tRNA* promoter has regulatory features and sequence elements in  
 common with the *tRNA* p1 promoter family of *Escherichia coli*.";  
 RL J. Bacteriol. 173:1757-1764(1991).  
 RN [4]

RP SEQUENCE OF 456-614 FROM N.A.  
 RC STRAIN=RD020;  
 RX MEDLINE=93106943; PubMed=8093236;  
 RA Dougherty T.J., Thanassi J.A., Pucci M.J.;  
 RT "The Escherichia coli mutant requiring D-glutamic acid is the result  
 of mutations in two distinct genetic loci";  
 RL J. Bacteriol. 175:111-116 (1993).  
 RN [5]  
 RP SEQUENCE OF 21-32.  
 RC STRAIN=K12 / EWG2.  
 RX MEDLINE=97443975; PubMed=9238646;  
 RA Link A.J., Robison K., Church G.M.;  
 RT "Comparing the predicted and observed properties of proteins encoded  
 in the genome of Escherichia coli K-12";  
 RL Electrophoresis 18:1259-1313 (1997).  
 RN [6]  
 RP MUTAGENESIS OF TONB BOX.  
 RX MEDLINE=90078094; PubMed=2687240;  
 RA Gudmundsdottir A., Bell P.E., Lundrigan M.D., Bradbeer C.,  
 RA Kader R.J.;  
 RT "Point mutations in a conserved region (TonB box) of Escherichia coli  
 outer membrane protein Btub affect vitamin B12 transport.";  
 RL J. Bacteriol. 171:6526-6533 (1989)  
 CC -1- FUNCTION: COBALAMIN RECEPTOR PROTEIN. ALSO ACTS AS A RECEPTOR FOR  
 BACTERIOPHAGE BF23 AND IS NECESSARY FOR THE UPTAKE OF E COLICINS.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane.  
 CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M10112; AAA23524.1; -;  
 DR EMBL; U00006; AAC3072.1; -;  
 DR EMBL; AE000471; AAC76948.1; -;  
 DR EMBL; M57568; -; NOT ANNOTATED\_CDS.  
 DR EMBL; L14556; AAA23676.1; -;  
 DR PIR; A21892; QRECBT.  
 DR EcoGene; EG10126; btub.  
 DR InterPro; IPR000531; TonB boxC.  
 DR Pfam; PF00593; TonB boxC; 1.  
 DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; 1.  
 DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
 KW Outer membrane; Cobalt transport; Transport; TonB box; Signal;  
 KW Phage recognition; Receptor; Complete proteome.  
 FT CHAIN 1 20  
 FT SITE 21 614 VITAMIN B12 RECEPTOR.  
 FT SITE 25 33 TONB BOX.  
 FT SITE 597 614 TONB C-TERMINAL BOX.  
 FT MUTAGEN 28 28 L->P; INACTIVATE UPTAKE.  
 FT MUTAGEN 30 30 V->G; INACTIVATE UPTAKE.  
 FT CONFLICT 162 162 A -> G (IN REF. 1).  
 FT CONFLICT 377 377 A -> R (IN REF. 1).  
 SQ SEQUENCE 614 AA; 66407 MW; AB43C64A991F95 CRC64;  
 Query Match 6.2%; Score 126; DB 1; Length 614;  
 Best local Similarity 23.4%; Pred. No. 0.3;  
 Matches 67; Conservative 46; Mismatches 99; Indels 74; Gaps 13;

DB 146 VNNITTRDEPGTEISAGNSYQNVDSVTOOQGDKTRVTLGLDYANTHGVAVAYGN 205  
 QY 178 HGSKLSEGIQVGVQSVAGVSEN-----GNVIALRDVYRTTGAPANGRIAPBPA 231  
 DB 206 TGRQAQTDNDGF..LSKTVLGALEHNTDAMSGFRVGVGNKRNINAYYSFGSPL-----L 260  
 QY 232 QTDKODSKSLSVNTVNDVQLDKQNNINLALTHYNDKQDTPDYADYG 277  
 DB 261 DTRKLVYSQSWDAGLRNGELIKSQ-----LITSYSHKDVYDPHYG 302  
 RESULT 17  
 FCT ERMCH  
 ID FCT ERMCH STANDARD; PRT; 735 AA.  
 AC 047162;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE Ferrichyobactin receptor precursor.  
 GN FCT.  
 OS Erwinia chrysanthemi.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Pectobacterium.  
 OX NCBI\_TaxID=556;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=3937;  
 RX MEDLINE=96165286; PubMed=8576065;  
 RA Sauvage C., Franca T., Expert D.;  
 RT "Analysis of the Erwinia chrysanthemi ferrichyobactin receptor  
 gene: resemblance to the Escherichia coli fepA-fes bidirectional  
 promoter region and homology with hydroxamate receptors";  
 RL J. Bacteriol. 178:1227-1231 (1996).  
 CC -1- FUNCTION: INVOLVED IN THE INITIAL STEP OF IRON UPTAKE BY BINDING  
 CHRISOACTIN, AN IRON CHELATIN SIDEROPHORE THAT ALLOWS THE  
 BACTERIA TO EXTRACT IRON FROM THE ENVIRONMENT.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane.  
 CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X87967; CAA61205.1; -;  
 DR HSSP; P06971; 2FCP.  
 DR InterPro; IPR000531; TonB boxC.  
 DR Pfam; PF00593; TonB boxC; 1.  
 DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; FALSE\_NEG.  
 DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; 1.  
 KW Signal; Receptor; Transmembrane; Outer membrane; Iron transport;  
 KW Transport; TonB box.  
 FT CHAIN 1 38  
 FT SITE 39 735  
 FT SITE 45 52  
 FT SITE 718 735  
 FT DOMAIN 39 45 PERIPLASMIC (POTENTIAL).  
 FT DOMAIN 46 54 PERIPLASMIC (POTENTIAL).  
 FT DOMAIN 55 82 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 83 91 POTENTIAL.  
 FT DOMAIN 92 107 PERIPLASMIC (POTENTIAL).  
 FT TRANSMEM 108 116 POTENTIAL.  
 FT DOMAIN 117 124 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 125 133 POTENTIAL.  
 FT DOMAIN 134 159 PERIPLASMIC (POTENTIAL).  
 FT TRANSMEM 170 178 POTENTIAL.  
 FT DOMAIN 179 179 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 180 188 POTENTIAL.  
 FT DOMAIN 189 194 PERIPLASMIC (POTENTIAL).  
 FT TRANSMEM 195 203 POTENTIAL.



QY 288 PSNNAIKGSLSEQ 301  
DB 247 PSNNVEENELNEQ 260

RESULT 19  
FOXA\_SALTY STANDARD; PRT; 702 AA.  
AC Q56145;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DE Ferrioxamine B receptor precursor.  
GN FOXA OR STM0364.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OK NCBI\_TaxId=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RA MEDLINE=21534948; PubMed=11677699;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Lettelle P.,  
Courtney L., Porciliak S., Ali J., Dante M., Du F., Hou S., Layman D.,  
Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,  
Waterston R., Wilson R.K.;  
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium  
LT2.";  
RT Nature 413:852-856 (2001).  
RN [2]  
RP SEQUENCE OF 1-374 FROM N.A.  
RC STRAIN=SL1344;  
RA MEDLINE=99203118; PubMed=10103258;  
RA Kingsley R.A., Ratsch R., Rabsch W., Ketley J.M., Tsolis R.M.,  
Everest P., Dougan A.J., Baumber A.J., Roberts M., Williams P.H.;  
RT "Ferrioxamine-mediated iron(III) utilization by Salmonella  
enterica.";  
RT Appl. Environ. Microbiol. 65:1610-1618 (1999).  
RN [3]  
RP SEQUENCE OF 1-244 FROM N.A.  
RC STRAIN=ATCC 14028;  
RA MEDLINE=95370139; PubMed=7642488;  
RA Tsolis R.M., Baumber A.J., Stojiljkovic I., Heffron F.;  
RT "Purification of Salmonella typhimurium: identification of new iron-  
regulated genes.";  
RT J. Bacteriol. 177:4628-4637 (1995).  
RN [4]  
RP FUNCTION: FERRIOXAMINE BINDING AND UPTAKE, IN ASSOCIATION WITH THE  
TONB PROTEIN (BY SIMILARITY). MAY PLAY A ROLE IN INTESTINAL  
COLONIZATION.  
CC -1- SUBCELLULAR LOCATION: Outer membrane.  
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC EMBL; AE008712; AAL19318.1;  
CC EMBL; AF060876; AAC15464.1;  
CC EMBL; U62282; AAB04552.1;  
CC HSSP; P06971, 1BY5  
DR StyGene; SG10646; FOXA.  
DR InterPro; IPR000531; TONB\_boxC.  
DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; FALSE NEG.  
DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; FALSE NEG.  
KW Signal; Receptor; Transmembrane; Outer membrane; Iron transport;  
KW Transport; TonB box; Complete proteome.  
FT SIGNAL 1 30 POTENTIAL.

FT CHAIN 31 702 FERRIOXAMINE B RECEPTOR.  
FT CONFLICT 246 249 MAYP -> SPYL (IN REF. 2).  
FT CONFLICT 255 255 G -> R (IN REF. 2).  
FT CONFLICT 279 279 F -> S (IN REF. 2).  
FT CONFLICT 352 352 A -> P (IN REF. 2).  
FT CONFLICT 368 374 HRLVGI -> SSSAGD (IN REF. 2).  
SQ SEQUENCE 702 AA; 77685 MW; 7E2CBA05B09E860D CRC64;  
Query Match 5.9%; Score 120.5; DB 1; Length 702;  
Best Local Similarity 20.2%; Pred. No. 0.85;  
Matches 80; Conservative 57; Mismatches 159; Indels 101; Gaps 16;  
QY 20 TQGLYQAPNESLPTVLEBPVIT-IDKSGMALANRITQWPHTKYIYERQIQEO-ATGS 76  
DB TPLPLAQ-----ETTNDRTVIVTSFVQSGATKATPDIETPQSVSITITQCFEQATSV 79  
QY 77 RQL-----ADVMAQLIPSLGVSSGTTSPFGQTMHGRQVQLNGVPLTGRSD 123  
DB RQAVSYTPGVYSNQIGASNRFDVILRGFSDSLDN-----VVLDDGLKMMGDDTV 128  
QY 124 ISRQLNSINPNOVARIEVLGATSI-YGSGATGGLINITYKSDLBEQFETRIGVAGSKL 182  
DB 129 SHSSL-VDPWFLEIDIEVVRGPASVLYGRSSPGIVALTSRKAPDAGBVKL--FAGNN 185  
QY 183 SSEGIGYQGVQSVAGVSENGVYLR-----DVEDYRT----- 215  
DB 186 NQGAAPDV---TGPLDNERVVAARLSGMTRYADSGFTLKEERYALMPSLWIRIDRTR 242  
QY 216 -----GAPDA-----NGKRIA---DEPAQTDYQDSKSLSVNTNVD 249  
DB 243 LPLMAYPHRDPEGSHSGPYQGTVPYNGKISNTFFEGEDDYDYDRRENNVGVNIIEH 302  
QY 250 QLDDKQINILALTHVNDKQDTD--VAPDGNLAVLFGKPSLNAIKGSLSRPRTTTS 307  
DB 303 LPDNGSVQKRLRYHTKTYTLAQVYAGLNLATLRKGSQSGEKMSAIALDNLQDGSVD 362  
QY 308 TFINYNH---HDDLWGNTINTNAYRERKGRFPYFVA 341  
DB 363 TGAINRLLVGVIDYQDRSNHTTGY---GAPPEIDA 395  
RESULT 20  
HPUB NEIMA STANDARD; PRT; 810 AA.  
AC Q9JWA2;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE Hemoglobin-haptoglobin utilization protein B precursor.  
GN HPUB OR NMA0474.  
OS Neisseria meningitidis (serogroup A).  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OK NCBI\_TaxId=656599;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Z2491 / Serotype 4A;  
RA MEDLINE=20222556; PubMed=10761919;  
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,  
Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holtroyd S.,  
Jagals K., Leather S., Moule S., Mungall K., Quail M.A.,  
Rajadhead S., Rutherford K.M., Skjold M., Skelton J.,  
Whitehead S., Spratt B.G., Barrett B.G.;  
RT "Complete DNA sequence of a serogroup A strain of Neisseria  
meningitidis Z2491.";  
RN [2]  
RL Neutrigidids Z2491.  
CC -1- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE  
HEMOGLOBIN/HAPTOGLOBIN COMPLEX AND IS REQUIRED FOR HEME UPTAKE (BY  
SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Outer membrane (potential).  
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; AL162753; CAB83769.1; --  
DR InterPro; IPR000531; TonB\_boxC.  
DR Pfam; PF00593; TonB\_boxC; 1.  
DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; FALSE\_NEG.  
DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
KW Outer membrane; Transport; TonB\_box; Signal; Receptor;  
KW Complete proteome.  
FT SIGNAL 1 22 POTENTIAL.  
FT CHAIN 23 810 HEMOGLOBIN-HAPTOGLOBIN UTILIZATION  
FT FT PROTEIN B.  
FT SITE 793 810 TONB C-TERMINAL BOX.  
FT SEQUENCE 810 AA; 90570 MW; D38DEIDCA3CA5A6E CRC64;

Query Match 5.9%; Score 120; DB 1; Length 810;  
Best Local Similarity 21.8%; Pred. No. 1.1;  
Matches 75; Conservative 50; Mismatches 131; Indels 88; Gaps 20;

Qy 13 PLLSVAVTQOLYAQNESLPTVELEPVVITIDKSGMALANRITQM-PHTTKVIYEEQIQE 71  
Db 7 PVLAALAAQAAP-----PAPAADPA-----PQSAQTL-NEITVTGTHKTQKGEKIRR 54  
Qy 72 QATGSRQLADV--MAQLIPSLG-VSSGTTSNFGQTMHG---RQVQFLNGVPLTISR--- 122  
Db 55 KTLDKLLVNDHDLVRYDPGIVSVEGGRAGNGFTIRGVKDRVAINVDGLAQAEBSSE 114  
Qy 123 -----DISRLQNSINPNQVARIIEVLGATSI-YGSGATGGLINIVTK--SDLSEE 169  
Db 115 AFOELFGAYGNPNANRNTSEPNFSEVITTKGADSLKSGSGALGAVNYQTKSADSYVSE 174  
Qy 170 QPETRIGVHG-----SKLSSEG-----IGY--QVGQSVAGVSENGVRLRLD 209  
Db 175 DKPYHLGKGGSVGKNSOKFSSITAGRLFGLDLALLVYTRFEKTKNRSTEGDIEIKND 234  
Qy 210 ---VDYRTTGG-----AFDANG-KRIAPPAQ--TDKQDSKLSLVNTN---VDWQLDDKQ 255  
Db 235 GVVYNPDTGGPSKYLTVYATGVARSQDPQEWNKSTLFLKLYNFNDQNRIGWIFEDSR 294  
Qy 256 -----NINLALHYNDKQDITY-----APDYGRL 280  
Db 295 TDRFTNELSNLWGTGTTSAATGDRHRQDVSYRRRSRSGVEYKNEL 338

RESULT 21  
FLAA\_HELPF STANDARD; PRT; 512 AA.  
ID FLAA\_HELPF  
AC O9XB38;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Flagellin A.  
GN FLAA.  
OS Helicobacter felis.  
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
OC Helicobacter.  
OX NCBI\_TaxID=214;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CS1;  
RX MEDLINE=99340214; PubMed=10411751;  
RA Josephans C., Ferrero R.L., Labigne A., Suerbaum S.;  
RT "Cloning and allelic exchange mutagenesis of two flagellin genes of  
RT Helicobacter felis";  
RL Mol. Microbiol. 33:350-362 (1999).  
CC -!- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO  
CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA. IMPORTANT FOR MOTILITY  
CC AND VIRULENCE.

CC -!- SUBUNIT: HETEROPOLYMER OF FLAA AND FLAB. FLAB IS LOCATED  
CC PROXIMAL TO THE HOOK WHILE THE REMAINDER OF THE FILAMENT IS  
CC COMPOSED OF THE PREDOMINANT FLAA.  
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.  
CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; Y11601; CAB46858.1; --  
DR InterPro; IPR001492; FlagellinN.  
DR InterPro; IPR001029; FlagellinC.  
DR Pfam; PF00669; FlagellinN; 1.  
DR Pfam; PF00700; FlagellinC; 1.  
DR PRINTS; PR00207; FLAGELLIN.  
DR ProDom; PD000316; Flagellin\_C; 1.  
KW Flagella.  
FT INIT\_MET 0 0 BY SIMILARITY.  
SQ SEQUENCE 512 AA; 53397 MW; B40BB6BF272C256A CRC64;

Query Match 5.8%; Score 118.5; DB 1; Length 512;  
Best Local Similarity 22.2%; Pred. No. 0.75; Indels 109; Gaps 20;  
Matches 88; Conservative 60; Mismatches 140;

Qy 22 QLYAQPNESLPTVELEPVVITIDKSGMALANRITQM-PHTTKVIYEEQIQE-QATGSRQL- 79  
Db 62 QAIANTNDGMIQI-----ADK---AMDEQI-KILDTIKVKATQAQDQGTQSRKAL 111  
Qy 80 -ADVMAQLIPSLGVSSGTTSNFGQTM-----HGRQVQF----- 111  
Db 112 QADIV-RLIQSLDNLIGNTTSYNGQTLGSGAPSNKEFQVGAYSNETIKASIGSATSDKIGQ 170  
Qy 112 --LLNGVPLTGSRLDIS---RQLNSINPNQVARIIEVLGATSIYG-----SGATG-- 155  
Db 171 VKITTKGNITAGSEVALTFKQDGVHDVLSVSKISTASAGTGLGLVLAIEVINKNSNATGVR 230  
Qy 156 GLINIVTKSLDEEEQFETRIQVHGSKLSSEGI-QVGQSVAGVSEN---GNVLRLD 211  
Db 231 ATANVITTS-----SAIKSGSLSLTVNGEIG-NILGIKNDSDGLVLAALNAV 280  
Qy 212 YRTTGAFFDANGKRIAPPAQTDQ-----DSKLSV---NTNVWQJDDKQINILA 260  
Db 281 TAQTG-----VEAYTDSVGRNLNRSIDGRGINIKANSTNVDSQAS-----A 321  
Qy 261 LTHYNDKQDITYADYGNRLAVLFGEKPSLNAIKLSLSEQPKTKSTFNINHHDDLWG 320  
Db 322 LTTLGGQDITRGSTNFGRLSLVRQDARDILVSGANVSASAGYAAIGP-----AKG 373  
Qy 321 NTINTNAYRREKGRFPFVAPFSAIKALPIQSMNL 357  
Db 374 TTANTVNLRLDVLGEFNQAVRSAGSANYNKTASENL 410

RESULT 22  
TB12\_NEIMB STANDARD; PRT; 908 AA.  
ID TB12\_NEIMB  
AC Q05987;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Transferrin-binding protein 1 precursor.  
GN TBPI.  
OS Neisseria meningitidis (serogroup B).  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=491;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-42.  
RC STRAIN=CCUG 37603 / B16B6 / Serogroup B / Serotype 2a;  
RX MEDLINE=93345825; PubMed=8344530;



RA Legrain M., Mazarin V., Irwin S.W., Bouchon B., Quentin-Millet M.-J.,  
 RA Jacobs E., Schryvers A.B.;  
 RT "Cloning and characterization of *Neisseria meningitidis* genes  
 RT encoding the transferrin-binding proteins Tbp1 and Tbp2.",  
 RL Gene 130:73-80(1993).  
 RN [2]  
 RP SEQUENCE OF 25-45.  
 RC STRAIN=CCUG 37603 / B16B6 / Serogroup B / Serotype 2a;  
 RX MEDLINE=93307625; PubMed=8319886;  
 RA Giffiths E., Stevenson P., Byfield P., Ala'Aldeen D.A.,  
 RA Borriello S.P., Holland J., Parsons T., Williams P.;  
 RT "Antigenic relationships of transferrin-binding proteins from  
 RT *Neisseria meningitidis*, *N. gonorrhoeae* and *Haemophilus influenzae*:  
 RT cross-reactivity of antibodies to NH2-terminal peptides.",  
 RL FEMS Microbiol. Lett. 109:85-91(1993).  
 CC -1- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR  
 CC TRANSFERRIN UTILIZATION.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane.  
 CC -1- INDUCTION: BY IRON STARVATION.  
 CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; Z15129; CA78831.1; -;  
 DR PIR; JN0819; JN0819.  
 DR PIR; S33154; S33154.  
 DR InterPro: IPR00531; TONB boxC.  
 DR Pfam: PF00593; TONB boxC.  
 DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; 1.  
 DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
 DR Outer membrane; Receptor; Signal; TONB box.  
 FT SIGNAL 1 24  
 FT CHAIN 25 908 TRANSFERRIN-BINDING PROTEIN 1.  
 FT SITE 38 45 TONB BOX.  
 FT SITE 891 908 TONB C-TERMINAL BOX.  
 SQ SEQUENCE 908 AA; 101583 MW; FE2FP4974CAC3C31 CRC64;  
 Query Match 5.8%; Score 118; DB 1; Length 908;  
 Best Local Similarity 20.6%; Pred. No. 1.8;  
 Matches 89; Conservative 59; Mismatches 162; Indels 122; Gaps 22;  
 QY 1 MRHSYFQWLSPLLSAVTQQLVAQ--PNISLPVLEPVVITIDKSGMALANRITQMP 58  
 DB 1 MOOHLFR-LNITLCISMTALPVYAEVQAEQAEKQDITQVAKKQKTRRDNEVTGIG 59  
 QY 59 HTTYVYEQVQEQATSGROLADWAQQLPSIG-VSSGTSNFGQTMG--RQVQPLIN 114  
 DB 60 KLIVR-SSDTLSKEQVLNIRDL--TRYDPGLAVVEQGRGSSGISIKGMDKNRVSITVD 114  
 QY 115 GV-----PLTGSRDI--SRQINSINPQVARIETLSGA-TSIVSGATGLINIV 161  
 DB 115 GVSQIGSTAAALGCTTAGSSGAINIEHYENVAVAEISKGSNSSEYENGALAGSVAFQ 174  
 QY 162 TK--SDL--EEDEPFRIGVHSGKLSSEIGVQVQSVAGVSENGVNLAR----- 207  
 DB 175 TKTAADITIGEGKW---GIQ-SKTVAYSGKHALLQSLALGRSGAEMALITYTKRGRE 229  
 QY 208 -----LDVDYRTGAF-----DANGKRLA 227  
 DB 230 IAHAKDAGVQSFNRVLDEKKGSGQYRYFVEECSHNGYAACKKRLKEDASVKDER 289  
 QY 228 PEPAQTDKQDKSLSVNT-----NVDWOLDKQONIALTHYNDKQDT----- 270  
 DB 290 KTVSTQDVTGSRNRLANLEYGSSQWMLRPGMHLNDRHYGVAVLERQTQPTTRMTYPA 349  
 QY 271 ----DYP-----DYGRLAVLF--GEKSLNMI-KGLSLSEQPKTKSTFNINY 313

DB 350 YFTSEDPVSGLSKGLKGYKSGDNKAERLFEVQSGESTLQIGYGVFDEHHTKNRYGVEY 409  
 QY 314 -HHD--DLWGN 321  
 DB 410 YFNADKDTWAD 421  
 RESULT 23  
 FYUA\_YERPE STANDARD; PRT; 673 AA.  
 AC P46359;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Pesticin receptor precursor (IRPC).  
 GN FYUA OR PSN OR YP01906.  
 OS *Yersinia pestis*.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC *Yersinia*.  
 OX NCBI\_TaxId=632;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=6/69;  
 RA Rakin A., Heesemann J.;  
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=KIT6;  
 RX MEDLINE=95204350; PubMed=7896707;  
 RA Fetherston J.D., Lillard J.W., Perry R.D.;  
 RT "Analysis of the pesticin receptor from *Yersinia pestis*: role in  
 RT iron-deficient growth and possible regulation by its siderophore.",  
 RL J. Bacteriol. 177:1824-1833(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=6/69;  
 RA Buchrieser C., Rusniok C., Couve E., Frangeul L., Billault A.,  
 RA Kunst F., Carniel E., Glaeser P.;  
 RT "DNA sequence of the 102 kbaes unstable region of *Yersinia pestis*."  
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CO-92 / Biovar Orientalis;  
 RX MEDLINE=21470413; PubMed=11586360;  
 RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,  
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,  
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Taraga A.M.,  
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,  
 RA Feltham T., Hamlin N., Holroyd S., Jagsis K., Karlyshev A.V.,  
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,  
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;  
 RT "Genome sequence of *Yersinia pestis*, the causative agent of plague.",  
 RL Nature 413:523-527(2001).  
 CC -1- FUNCTION: RECEPTOR FOR THE BACTERIOCIN PESTICIN AND FOR THE  
 CC SIDEROPHORE YERSINIACTIN.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane.  
 CC -1- INDUCTION: BY IRON STARVATION. IRON REGULATION MEDIATED THROUGH  
 CC THE FUR PROTEIN.  
 CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; Z15104; CA84487.1; -;  
 DR EMBL; U05530; AAA6906.1; -;  
 DR EMBL; AL031866; CAA21395.1; -;  
 DR EMBL; AJ414150; CAC90722.1; -;  
 DR HSSP; P05825; 1PEP.

DR	InterPro; IPR000531; TonB_boxC.	RP	SEQUENCE FROM N.A.
DR	Pfam; PF00593; TonB_boxC_1.	RC	STRAIN=WCS358;
DR	PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE NEG.	RX	MEDLINE=93316856; PubMed=8392140;
DR	PROSITE; PS01156; TONB_DEPENDENT_REC_2; FALSE NEG.	RA	Koster M., van de Vossen J., Leong J., Weisbeek P.J.;
KW	Signal; Receptor; Transmembrane; Outer membrane; Iron transport;	RT	"Identification and characterization of the pubB gene encoding an
KW	Transposon; TonB box; Complete proteome.	RT	inducible ferric-pseudobactin receptor of Pseudomonas putida
FT	SIGNAL 1 22 POTENTIAL.	RT	WCS358";
FT	CHAIN 23 673 PESTICIN RECEPTOR.	RL	Mol. Microbiol. 8:591-601(1993).
FT	SITE 30 37 TONB BOX.	CC	-I- FUNCTION: SPECIFIC RECEPTOR FOR THE SIDEROPHORES FERRIC
FT	SITE 657 673 TONB C-TERMINAL BOX.	CC	PSEUDOBACTIN BN8 AND FERRIC PSEUDOBACTIN BN7, IRON CHELATING
FT	CONFLICT 484 484 G -> D (IN REF. 1).	CC	MOLECULES THAT ALLOW THE ORGANISM TO EXTRACT IRON FROM THE
FT	CONFLICT 514 514 R -> G (IN REF. 1).	CC	ENVIRONMENT, SPECIALLY UNDER IRON-RESTRICTED CONDITIONS.
SQ	SEQUENCE 673 AA; 73782 MW; 9C39B6010EBCEB2C CRC64;	CC	-I- SUBCELLULAR LOCATION: Outer membrane.
		CC	-I- INDUCTION: BY IRON LIMITATION IN ADDITION TO THE PRESENCE OF ONE
		CC	OF THE TWO COGNATE PSEUDOBACTINS BN8 OR BN7.
		CC	-I- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
		CC	-----
		CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
		CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
		CC	the European Bioinformatics Institute. There are no restrictions on its
		CC	use by non-profit institutions as long as its content is in no way
		CC	modified and this statement is not removed. Usage by and for commercial
		CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
		CC	or send an email to license@isb-sib.ch).
		CC	-----
DR	EMBL; X73598; CAAS1995.1; -	DR	EMBL; X73598; CAAS1995.1; -
DR	PIR; S32899; S32899.	DR	PIR; S32899; S32899.
DR	InterPro; IPR000531; TonB_boxC.	DR	InterPro; IPR000531; TonB_boxC.
DR	Pfam; PF00593; TonB_boxC_1.	DR	Pfam; PF00593; TonB_boxC_1.
DR	PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE NEG.	DR	PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE NEG.
DR	PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.	DR	PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW	Outer membrane; Iron transport; Transport; Signal; Receptor.	KW	Outer membrane; Iron transport; Transport; Signal; Receptor.
FT	SIGNAL 1 45 POTENTIAL.	FT	SIGNAL 1 45 POTENTIAL.
FT	CHAIN 46 809 FERRIC-PSEUDOBACTIN BN7/BN8 RECEPTOR.	FT	CHAIN 46 809 FERRIC-PSEUDOBACTIN BN7/BN8 RECEPTOR.
FT	DOMAIN 156 160 POLY-SER.	FT	DOMAIN 156 160 POLY-SER.
FT	SITE 792 809 TONB C-TERMINAL BOX.	FT	SITE 792 809 TONB C-TERMINAL BOX.
SQ	SEQUENCE 809 AA; 88389 MW; 0B339F6E788ABCD CRC64;	SQ	SEQUENCE 809 AA; 88389 MW; 0B339F6E788ABCD CRC64;
	Query Match 5.7%; Score 117; DB 1; Length 809;		Query Match 5.7%; Score 117; DB 1; Length 809;
	Best Local Similarity 21.3%; Pred. No. 1.8;		Best Local Similarity 21.3%; Pred. No. 1.8;
	Matches 82; Conservative 69; Mismatches 152; Indels 82; Gaps 21;		Matches 82; Conservative 69; Mismatches 152; Indels 82; Gaps 21;
QY	26 QPNESLPTVELEPVI-----TIDKSGM-----ALANRIT-----QMPHTTKYIEEQ 68	QY	26 QPNESLPTVELEPVI-----TIDKSGM-----ALANRIT-----QMPHTTKYIEEQ 68
Db	122 QASATGALELSAVSISGKAPGSTTEGTLTYTSSSSSTRNLNLTPTETPQSLVTMTQR 181	Db	122 QASATGALELSAVSISGKAPGSTTEGTLTYTSSSSSTRNLNLTPTETPQSLVTMTQR 181
QY	69 IOQATGSRQLADVMAQLIPSL-----GVSSGTTNFGOTMGRQVQ-FLNGVPLTGS 122	QY	69 IOQATGSRQLADVMAQLIPSL-----GVSSGTTNFGOTMGRQVQ-FLNGVPLTGS 122
Db	182 LDQRL--TNLTDL--EATPGITVVDGLGSESDSYWSR---GFAIQNYEVDGP----- 230	Db	182 LDQRL--TNLTDL--EATPGITVVDGLGSESDSYWSR---GFAIQNYEVDGP----- 230
QY	123 DISRQLNSINPQVA--RIEVLGATS--YIGSGATGGLINIVTKSDLEEEQFETRIGVHG 179	QY	123 DISRQLNSINPQVA--RIEVLGATS--YIGSGATGGLINIVTKSDLEEEQFETRIGVHG 179
Db	231 -TSTRLDNYSQSMAFDRVEIVRGATGLISGMGNPSATINLIRKPTAEAQ----- 280	Db	231 -TSTRLDNYSQSMAFDRVEIVRGATGLISGMGNPSATINLIRKPTAEAQ----- 280
QY	180 SKLSSEGIG---YVQGSVAG--VSENGVNLRLDVYRTGGAFDANGKRIAPETADTK 235	QY	180 SKLSSEGIG---YVQGSVAG--VSENGVNLRLDVYRTGGAFDANGKRIAPETADTK 235
Db	281 ASITGEAGNMDRYGTGTFDVSGLPTETNGIRGFVADYKT-----EKAWIDR 326	Db	281 ASITGEAGNMDRYGTGTFDVSGLPTETNGIRGFVADYKT-----EKAWIDR 326
QY	236 QDSKSLSVNTNVDQLDDKQINLALTHYNDKQTDYAPDYGNNRLAVLGE-----KPSLN 291	QY	236 QDSKSLSVNTNVDQLDDKQINLALTHYNDKQTDYAPDYGNNRLAVLGE-----KPSLN 291
Db	327 YNOQSOLMYGITEFDLSEDTLLTVGFSYLRSDIDSPLRSGLPTRFST--GERTNLKRSLN 384	Db	327 YNOQSOLMYGITEFDLSEDTLLTVGFSYLRSDIDSPLRSGLPTRFST--GERTNLKRSLN 384
QY	292 AIKGLSLEQPKTKSTFPNHYHDDLWGNNTN--TNAYYRREK-----GRPYFPFVAPFS 344	QY	292 AIKGLSLEQPKTKSTFPNHYHDDLWGNNTN--TNAYYRREK-----GRPYFPFVAPFS 344
Db	385 AAPDWSVNDHEQTSFPT-SIEQQLGNGWSGKIELTHAENKFDLNFAMGELNPDGSLG 443	Db	385 AAPDWSVNDHEQTSFPT-SIEQQLGNGWSGKIELTHAENKFDLNFAMGELNPDGSLG 443
QY	345 IAKALPILOSMNLPSATLDAYTKAP 369	QY	345 IAKALPILOSMNLPSATLDAYTKAP 369
Db	444 ---QLPVRFSGTPTQDNLDLYATGP 465	Db	444 ---QLPVRFSGTPTQDNLDLYATGP 465
	RESULT 25		RESULT 25
	LBPA_NEIMA		LBPA_NEIMA

ID LBPA\_NEIMA STANDARD; PRT; 944 AA.  
AC Q9JTK4;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Lactoferrin binding protein A precursor.  
GN LBPA OR NMA1739.  
OS Neisseria meningitidis (serogroup A).  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI\_TaxId=65699;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=22491 / Serogroup A / Serotype 4A;  
RC MEDLINS=2022556; PubMed=10761919;  
RA Parhill J., Achman M., James K.D., Bentley S.D., Churcher C.,  
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,  
RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,  
RA Jagsis K., Leather S., Moule S., Mungall K., Quail M.A.,  
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,  
RA Whitehead S., Spratt B.G., Barrall B.G.;  
RT "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491."  
RT Nature 404:502-506 (2000).  
RL Nature 404:502-506 (2000).  
CC - FUNCTION: UNKNOWN. MAY BE AN IRON-SIDEROPHORE RECEPTOR.  
CC - SUBCELLULAR LOCATION: Outer membrane.  
CC - SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
CC - This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; AL162757; CAB84967.1; -  
DR InterPro; IPR000531; TONB\_boxC.  
DR Pfam; PF00593; TONB\_boxC; 1.  
DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; FALSE\_NEG.  
DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
KM Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor; Complete proteome.  
KW  
FT SIGNAL 1 27 POTENTIAL.  
FT CHAIN 28 944 LACTOFERRIN BINDING PROTEIN A.  
FT SITE 827 944 TONB\_C-TERMINAL\_BOX.  
SQ SEQUENCE 944 AA; 105732 MW; CE06B8192E74AE3E CRC64;  
Query Match 5.7%; Score 116.5; DB 1; Length 944;  
Best Local Similarity 19.3%; Pred. No. 2.4;  
Matches 95; Conservative 59; Mismatches 146; Indels 191; Gaps 21;  
QY 1 MNRSHYFQWLSLPLSLVAVYQQLYA-----QPNESLPTVLEBPVITITDSSGAL--ANR 53  
DB 1 MNKRHPS-LTLTALAIAAIPSYANPETAAADAAQSGLKEITYRAAAVGRSRKATG 59  
QY 54 ITQMPTTKVYEEOIQEQTGSRQLADVAQGLIPSLG-VSSGTTSNFGQTMHG--ROY 109  
DB 60 LCKIVTSETLNEQV-----LGIRDL-----TRYDPGAVAVGQNGASGYSINGVDKRY 111  
QY 110 QPLANGVPLT-----GSRDISRQLNSINPNQVARIETLSGA-TSIYSGATG 156  
DB 112 AVASVDVAQIQAFYVQSLSGYGGSGGSGAINEIYEENISTVEIDKAGASDHGSGALGG 171  
QY 157 LNIIVK--SDL-----BEQFETRIYGHGKSLSSKGI----- 187  
DB 172 AVAFRTKEADLLSDGKSWGIOAKTAVGSKNRQFMKSLAGFSKDKGWEGLLITERQGR 231  
QY 188 -----GYGVG-----OSVAGSENGVLT-- 205  
DB 232 TRHGDIAQVVEIGIRLDAFRQTYDIQKONKKAAYFLAGEBERPKRVATLAENGNTLKN 291  
QY 206 -----ARLDVDYRTTGAFAFGKRIAPPEAQTD 234

DB 292 QLNRRVEERKKNOPPLNBEEMAVEEAQAHENLSAOSYTGCG-----RILPDP----- 340  
QY 235 KODSKSLSVNTVMDQL-----DBKONINLALTHYNDKQDTPYAPYGNLAVL 283  
DB 341 -MDYRSGSLALGTRFGRGRHYVGVPEFTKORYIDRMTEQYQYGTDAKPSKSGYV 399  
QY 284 FGE-----KPSLNAIKGLSLSEQPKTTSSTFNINVHDDLNGNTINTNAY 328  
DB 400 DGNDFRDLGLYFVNIEEMKGDITNLVXGICL-----KYSRTKF-IDENH----- 441  
QY 329 YRREKGRFYPP 339  
DB 442 RRRMGLLYRY 452  
RESULT 26  
HGBA\_HAEDU  
ID HGBA\_HAEDU STANDARD; PRT; 972 AA.  
AC Q47952;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemoglobin binding protein precursor.  
GN HGBA.  
OS Haemophilus ducreyi.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxId=730;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=35000;  
RC MEDLINE=95286263; PubMed=7768598;  
RA Elkins C., Chen C.J., Thomas C.E.;  
RT "Characterization of the hgba locus encoding a hemoglobin receptor from Haemophilus ducreyi."  
RT J. Infect. Immun. 63:2194-2200 (1995).  
RL - FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OF THE HOST AND IS REQUIRED FOR HEM UPTAKE.  
CC - SUBCELLULAR LOCATION: Outer membrane.  
CC - SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY; HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; U17281; AAA75311.1; -  
DR InterPro; IPR000531; TONB\_boxC.  
DR Pfam; PF00593; TONB\_boxC; 1.  
DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; FALSE\_NEG.  
DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; FALSE\_NEG.  
KM Outer membrane; Transport; TonB box; Multigene family; Signal; Receptor.  
KW  
FT SIGNAL 1 22 POTENTIAL.  
FT CHAIN 23 972 HEMOGLOBIN BINDING PROTEIN.  
FT SITE 32 39 TONB\_BOX.  
FT SITE 955 972 TONB\_C-TERMINAL\_BOX.  
SQ SEQUENCE 972 AA; 110937 MW; DEDF0P9F3575FDAF CRC64;  
Query Match 5.7%; Score 116.5; DB 1; Length 972;  
Best Local Similarity 19.2%; Pred. No. 2.5;  
Matches 67; Conservative 62; Mismatches 145; Indels 75; Gaps 16;  
QY 10 LSLPLSLVAVTQGLYAQPNESLPTVLEBPVITITDSSGALNRTITQMHTTKVYEEOI 69  
DB 9 IITCIIIGVYHT--VVAESH--MOTKELRTIVSSSEDSVANKN-VGEIKNSALSKQOV 63  
QY 70 QEQATGSRQLADVAQGLIPSLGVSAGTTSNFGQTMHG--ROYQPLANGV----- 117

Db 64 QDSRLVRYETGVTV-----VEKRGFGSSGYAIRGVNDENRVAVVVGHLQHAETISSQ 116  
Qy 118 -----LTGSRDLSRQLNSINPNQVARIIVLSGATSI-YGSGATGLINIVTKSLDEEQQE 172  
Db 117 FKELPEGYGNFNTRNGVEVNLQAVIOKADAIKRTGSLGCTVGFESK-DARDYLID 175  
Qy 173 TRIGVHGSKLSSEGGIGYQVQSGVAG-----VSENGNVLARLDV 210  
Db 176 -----KNYHFGYKTYGSADNQKLSHVTAAAGRYSDFDLLAVHTQRHGNELR--NY 223  
Qy 211 DYRTTGGAFDANGKRIAPAPQATDKQDSKLSVNTVNDWQDDKQINLALTHYNDKQDT 270  
Db 224 GYRHYDGSV-VKREKADPYKITKQSSL-----IKIGVQLNDRFTLG---YDSRNT 274  
Qy 271 DYAPYVGNRLAVLFGKPSLNAIKGLSLSEQPKTKSTNNINHHDDLW 319  
Db 275 SRGTOWNSAFTSYNG-GPELKVDRHTNDQSNRKNISFVYE-NFDTNDEF 321

RESULT 27  
FYUA\_YEREN  
ID FYUA\_YEREN STANDARD; PRT; 673 AA.  
AC P46360;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Pesticin receptor precursor (IRPC) (IPR65).  
GN FYUA.  
OS Yersinia enterocolitica.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Yersinia.  
OX NCBI\_TaxID=630;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 51871 / WA-314 / Serotype O:8;  
RX MEDLINE=95075311; PubMed=7984105;  
RA Rakin A., Saken E., Harmsen D., Heesemann J.;  
RT "The pesticin receptor of Yersinia enterocolitica: a novel virulence  
factor with dual function.";  
RL Mol. Microbiol. 13:253-263(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=123-75-1, 8081, YE737, and 1209-79;  
RA Rakin A., Heesemann J.;  
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: RECEPTOR FOR THE BACTERIOCIN PESTICIN AND FOR THE  
SIDEROPHORE YERSINIACTIN.  
CC -!- SUBCELLULAR LOCATION: Outer membrane.  
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Z29675; CAAB2773.1; -;  
DR EMBL; Z35485; CAAB4620.1; -;  
DR EMBL; Z35486; CAAB4621.1; -;  
DR EMBL; Z35487; CAAB4622.1; -;  
DR EMBL; Z35496; CAAB4630.1; -;  
DR HSSP; P05825; 1FEP.  
DR InterPro; IPR000531; TonB\_boxC.  
DR Pfam; PF00593; TonB\_boxC\_1.  
DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; FALSE NEG.  
DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; FALSE NEG.  
KW Signal; Receptor; Transmembrane; Outer membrane; Iron transport;  
KW Transport; TonB box.  
FT SIGNAL 1 22 POTENTIAL.  
FT CHAIN 23 673 PESTICIN RECEPTOR.  
FT SITE 30 37 TONB BOX.

FT SITE 557 573 TONB C-TERMINAL BOX.  
FT VARIANT 362 362 N -> D (IN STRAINS 8081, YE737 AND 1209-  
79).  
SQ SEQUENCE 673 AA; 73682 MW; 1C20E0352FAD4DCC CRC64;  
Query Match 5.6%; Score 115.5; DB 1; Length 673;  
Best Local Similarity 19.5%; Pred. No. 1.8;  
Matches 110; Conservative 61; Mismatches 158; Indels 235; Gaps 26;  
Qy 10 LSLPLLSVAVTQOLYAQPNEISLPTVELEPVTITIDKSGMALANRITOMPHTTKVIYEQI 69  
Db 14 LLLPAIQAQT-----SQQDES--TLE--VTASKOSSRSASANNVS-----STVVSAPELS 60  
Qy 70 QOATGSGRLADVMAQLIPSLGVSSGTTSNF-----GQTMHGRQVQVQLNGVPLT 119  
Db 61 DAGVTAS-----DKLPRVLPGLNIENSGNMLFSTISLRGVSSAQDFYNPAVTLVGVGPQL 116  
Qy 120 GSRDISRQLNSINPNQVARIIVLSGAT-SIYSGATGGLINIVTKSD-----L 166  
Db 117 STNTQAL-----TDVQSVELLRPGQGLYKSAQGGIINIVTQOPDSTPRGYIEGVS 170  
Qy 167 EEEQETRIGVHGSKLSSEGIGY-----QVGOSVAGVSENGV 204  
Db 171 SRDSYRSKENLSGP--IQDGLLYGSVTLLRQVDDGDMINPATGSDDLGTRASI---GNV 225  
Qy 205 LARLDVD-----YRTTGGAP-----DANGKRI-----APEPAQ---TDKQ-- 236  
Db 226 KLRLAPDDQPMWEMGFAASRECTRATQDAYVGVNDIKGRKLSLSDGSPDPYMRCTDSOTL 285  
Qy 237 -----DSKLSLVN----- 244  
Db 286 SGKYTTDDWVFNLSAWQQHYRSTFPSSGLIVNMPQRMNQDVQLRAATLGDAITVDMV 345  
Qy 245 -----TNVDWQDDKQINLALTHYND 266  
Db 346 FGLYQNTREKLSAYNNPTWPLSYSTGYTTAETLAAVSDLTWHLTDFDGGVRFSDH 405  
Qy 267 KQDTYADPYGNRLAVLFGKPSLNAIKGLSLSEQPKTKSTFNINHHDDLNGNTINTN 326  
Db 406 KsSTQY---HGSMLGNPFQD-----QKSNDDQ---VLGQLSAGYMLTDDW-----R 446  
Qy 327 AYVREKGRFY-----PRVAPFSATAKALPI---LQSNLPSATLIDATYK 367  
Db 447 VTTRIAQG--YKPSGYNIVPTAGLDAKPFVAEKSINVELGTRYETADVTLQAATFYTHK 504  
Qy 368 APQARAYGV---LQESKAEVLG 387  
Db 505 DMQLYSGPVGWQTLNAGKADATG 528

RESULT 28  
YBIL\_ECOLI  
ID YBIL\_ECOLI STANDARD; PRT; 760 AA.  
AC P75780;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Probable tonB-dependent receptor ybil precursor.  
GN YBIL OR B0805.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474(1997).

[2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=K12;  
 RX MEDLINE=97061202; PubMed=8905232;  
 RA Oshima T., Aiba H., Baba T., Fujita K., Hayaishi K., Honjo A.,  
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,  
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,  
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,  
 RA Samped G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,  
 RA Yano M., Horuchi T.,  
 RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome  
 RT corresponding to the 12.7-28.0 min region on the linkage map.";  
 RL DNA Res. 3:137-155(1996).  
 [3]  
 RN IDENTIFICATION BY MASS SPECTROMETRY, AND PUTATIVE FUNCTION.  
 RX MEDLINE=20267815; PubMed=10806384;  
 RA Molloy M.P., Herbert B.R., Slade M.B., Rabioud T., Nouwens A.S.,  
 RA Williams K.L., Gooley A.A.,  
 RT "Proteomic analysis of the *Escherichia coli* outer membrane.";  
 RT Eur. J. Biochem. 267:2871-2881(2000).  
 CC -1- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT, THAT PARTICIPATES IN  
 CC IRON TRANSPORT.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane (Potential).  
 CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AE000182; AAC73892.1; -.  
 DR EMBL; D90717; BAA35471.1; -.  
 DR EMBL; D90718; BAA35476.1; -.  
 DR EcoGene; BG13317; ybll.  
 DR InterPro; IPR000531; TONB\_boxC.  
 DR Pfam; PF00593; TONB\_boxC; 1.  
 DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; 1.  
 DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; FALSE NEG.  
 KW Outer membrane; Iron transport; Transport; Tonb box; Signal; Receptor;  
 KW Complete proteome.  
 FT SIGNAL 1 33 POTENTIAL.  
 FT CHAIN 34 760 PROBABLE TONB-DEPENDENT RECEPTOR YBLL.  
 FT SITE 743 760 TONB C-TERMINAL BOX.  
 SQ SEQUENCE 760 AA; 81960 MW; 1B66B9C2EF68C7CA CRC64;  
 Query Match 5.6%; Score 114.5; DB 1; Length 760;  
 Best Local Similarity 22.1%; Pred. No. 2.4; Mismatches 155; Indels 69; Gaps 15;  
 Matches 75; Conservative 41; Mismatches 155; Indels 69; Gaps 15;  
 QY 14 LLSAVTQQLVAQPNESLPTVELEPVITIDKSGMALNRITQMPHTTVIYEEOIQEA 73  
 DB 44 LVEASTPSLVA-PQGSAPKFSRPVADT-----TETMTVISEQVIKDD- 86  
 QY 74 TGSRLQADYMAQLIPSLGV-----SSGTSNFGQTHGROVQGLNGVLLTSGRDI-SRQ 127  
 DB 87 -GATNTLTLAKN-VPGVGAFFAGENGSTTGDAIYRGADTS--NSIYIDGIRDIGSVS 141  
 QY 128 LNSINPQVARIELVLSGATSI-YGSGATGGLINIVTKSLDEEQQFTRIGVHSGKLSSEG 186  
 DB 142 RQTFNTEQV---EVIKGPGSTGGRSAFGOSINMISKORNDGIDASASI-QSAPFRFG 197  
 QY 187 ---IGYQVQSVAGVSENGVNLARLDVYRTTGAFDANGKRIAPPAQTDKQSSLSLV 243  
 DB 198 TLDVNQVIDTTLA-----VRLNVMEKKT-----HDAGRDKVKNERYGV 235  
 QY 244 NNNVMDQLDDKQINIALTHYNDKQDPT-----YAPDGNRLAVLFGKPKSLNMI 293  
 DB 236 ADSVAAGLGTANRLYLNTLVQHTNPDDGIPITGLPGVSAPAGTAAALNHSKVDTHNF 295  
 QY 294 KGLSLSEQPTTKSTFNINVYHDDLWGNNTINTNAVYRREK 333

DB 296 YGTD-SDYDSTDTATWRFPHDINDNTINTTRWSRVK 334  
 RESULT 29  
 ID FOXA\_YEREN STANDARD; PRT; 710 AA.  
 AC Q01674;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Ferrioxamine receptor precursor.  
 GN FOXA.  
 OS *Yersinia enterocolitica*.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC *Yersinia*.  
 OK NCBI\_TaxId=630;  
 [1]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=ATCC 51872 / WA-C / Serotype O:8;  
 RX MEDLINE=9234959; PubMed=1640832;  
 RA Baumann A.J., Hanke K.,  
 RT "Ferrioxamine uptake in *Yersinia enterocolitica*: characterization of  
 RT the receptor protein FoxA.";  
 RL Mol. Microbiol. 6:1309-1321(1992).  
 [2]  
 RN REVISIONS.  
 RC STRAIN=ATCC 51872 / WA-C / Serotype O:8;  
 RA Baumann A.J.;  
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: FERRIOXAMINE BINDING AND UPTAKE, IN ASSOCIATION WITH THE  
 CC TONB PROTEIN.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane.  
 CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X60447; CAA42975.1; -.  
 DR PIR; S22673; S22673.  
 DR HSSP; P06971; 10Q0.  
 DR InterPro; IPR000531; TONB\_boxC.  
 DR Pfam; PF00593; TONB\_boxC; 1.  
 DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; 1.  
 DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
 KW Signal; Receptor; Transmembrane; Outer membrane; Iron transport;  
 KW Transport; Tonb box.  
 FT SIGNAL 1 26  
 FT CHAIN 27 710 FERRIOXAMINE RECEPTOR.  
 FT SITE 28 35 TONB BOX.  
 FT SITE 693 710 TONB C-TERMINAL BOX.  
 FT TRANSMEM 29 37 POTENTIAL.  
 FT TRANSMEM 65 73 POTENTIAL.  
 FT TRANSMEM 91 99 POTENTIAL.  
 FT TRANSMEM 106 114 POTENTIAL.  
 FT TRANSMEM 137 145 POTENTIAL.  
 FT TRANSMEM 152 160 POTENTIAL.  
 FT TRANSMEM 180 188 POTENTIAL.  
 FT TRANSMEM 194 202 POTENTIAL.  
 FT TRANSMEM 208 216 POTENTIAL.  
 FT TRANSMEM 259 267 POTENTIAL.  
 FT TRANSMEM 271 279 POTENTIAL.  
 FT TRANSMEM 293 301 POTENTIAL.  
 FT TRANSMEM 309 317 POTENTIAL.  
 FT TRANSMEM 353 361 POTENTIAL.  
 FT TRANSMEM 370 378 POTENTIAL.  
 FT TRANSMEM 427 435 POTENTIAL.  
 FT TRANSMEM 443 451 POTENTIAL.



CC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 CC Helicobacter.  
 CC NCBI\_TaxID=210, 85963;  
 CC [1]  
 CC SEQUENCE FROM N.A., AND SEQUENCE OF 1-13.  
 CC STRAIN=NTCC 11637 / Isolate 898-1;  
 CC MEDLINE=93062020; PubMed=1435261;  
 CC Leying H., Sherbaum S., Geis G., Haas R.;  
 CC "Cloning and genetic characterization of a Helicobacter pylori  
 CC flagellin gene";  
 CC Mol. Microbiol. 6:2863-2874(1992).  
 CC [2]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=26695 / ATCC 700392;  
 CC MEDLINE=97394467; PubMed=9252185;  
 CC Tomb J.-F., White O., Kierlavage A.R., Clayton R.A., Sutton G.G.,  
 CC Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,  
 CC Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,  
 CC Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
 CC McElveney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,  
 CC Berg D.E., Goeyne J.D., Uutterback T.R., Peterson J.D., Kelley J.M.,  
 CC Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,  
 CC Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,  
 CC Venter J.C.;  
 CC "The complete genome sequence of the gastric pathogen Helicobacter  
 CC pylori";  
 CC Nature 388:539-547(1997).  
 CC [3]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=J99;  
 CC MEDLINE=99120557; PubMed=9923682;  
 CC Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown B.D., Doig P.C.,  
 CC Smith D.R., Noonan B., Guild B.C., deJonghe B.L., Carmel G.,  
 CC Tummino P.J., Carno A., Uria-Nickelsen M., Mills D.M., Ives C.,  
 CC Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.B., Voris G.P.,  
 CC Trust T.J.;  
 CC "Genomic sequence comparison of two unrelated isolates of the human  
 CC gastric pathogen Helicobacter pylori";  
 CC Nature 397:176-180(1999).  
 CC [4]  
 CC SEQUENCE OF 1-26, AND CHARACTERIZATION.  
 CC STRAIN=NTCC 11637 / Isolate 915, and NTCC 11637 / Isolate 5294;  
 CC MEDLINE=91123220; PubMed=1704004;  
 CC Kostorzynska M., Betts J.D., Austin J.W., Trust T.J.;  
 CC "Identification, characterization, and spatial localization of two  
 CC flagellin species in Helicobacter pylori flagella";  
 CC J. Bacteriol. 173:937-946(1991).  
 CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO  
 CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA. IMPORTANT FOR MOTILITY  
 CC AND VIRULENCE.  
 CC -1- SUBUNIT: HETEROPOLYMER OF FLAA AND FLAB. FLAB IS LOCATED  
 CC PROXIMAL TO THE HOOK WHILE THE REMAINDER OF THE FILAMENT IS  
 CC COMPOSED OF THE PREDOMINANT FLAA.  
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X60746; CAA43148.1; -  
 CC EMBL; AB000574; AAD07667.1; -  
 CC EMBL; AB001487; AAD06133.1; -  
 CC PIR; S26192; S26192.  
 CC TIGR; HP0601; -  
 CC InterPro; IPR001492; Flagellin\_N.  
 CC InterPro; IPR001029; Flagellin\_C.  
 CC Pfam; PF00669; Flagellin\_N; 1.  
 CC Pfam; PF00700; Flagellin\_C; 1.  
 CC PRINTS; PR00207; FLAGELLIN.

DR Prodom; PD000316; Flagellin\_C; 1.  
 KW Flagella; Complete proteome.  
 FT INIT MET 0  
 FT VARIANT 14 14 H -> G (IN ISOLATE 5294).  
 FT CONFLICT 18 18 A -> E (IN REF. 1).  
 SQ SEQUENCE 509 AA; 53153 MW; B802C0B7446833F7 CRC64;  
 Query Match 5.5%; Score 112.5; DB 1; Length 509;  
 Best Local Similarity 22.5%; Pred. No. 1.9;  
 Matches 78; Conservative 53; Mismatches 109; Indels 107; Gaps 18;  
 QY 60 TTKVIEBQIOE-QATGSRQ-LADVMAQLPSLGVSSGTTNFGQTEGRQV---QF--- 111  
 DB 90 TVKVVATQAAQDQTTESRKAIQSDIVRLIOGLDINIGTTTYNGGALLSGQFTNKEFOVG 149  
 QY 112 -----LNGVPLGSRDIS-----RQLNSINPNQVARIPLVLSGA 145  
 DB 150 AVSNQSIASIGSTTSKDIQGVRIATGALITASGDISLTFQVQDGVNDVTLSESVSSA 209  
 QY 146 TSIVSGATGGLINIVTKSDLEEQFETRIGV--HGSKLSEGLIGYQGVAGVSENGN 203  
 DB 210 ---CTG-IGVLAIVYNNK-----SKRTGVKAVASVITTSQVAVQSG-SLSNLTNGI 256  
 QY 204 VIARLDVYRTTGAFFDANGKRIAPPAQTDKQDSLSVNTVDWQLDKQNTNL----- 259  
 DB 257 HLGNI-ADIKKN---DSDGRVLAIVNVTSE-----TGVAVYDQKRLNLRSID 302  
 QY 260 -----ALTHYNDKOD-TDYAPDYGKRLAVLFGKSPSLNAIK----- 294  
 DB 303 GRGIEIKTDSVNGSPSLATMVNGQDLTKGSTNYG-RSLTRLDKSKSINVVASDSQHLG 361  
 QY 295 ---GLASEQPKTKSTFNINHHDDLMGNTINTNAVYRREKGRFY 337  
 DB 352 PTAIGFSGSQVAETTVNLADV-----TGNNAVYKASGANY 398  
 RESULT 32  
 HXCI\_HABIN STANDARD; PRT; 744 AA.  
 AC P44523;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Heme/hemopexin utilization protein C precursor.  
 GN HXUC OR H10113.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 CC NCBI\_TaxID=727;  
 CC [1]  
 CC SEQUENCE FROM N.A.  
 CC STRAIN=Rd / KW20 / ATCC 51907;  
 CC MEDLINE=95350630; PubMed=7542800;  
 CC Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 CC Kierlavage A.R., Sutton G., Tomb J.-F., Dougherty B.A., Merriek J.M.,  
 CC McKenney K., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
 CC Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,  
 CC Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 CC Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 CC Fine L.D., Fitzhugh J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
 CC Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 CC Venter J.C.;  
 CC "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 CC Rd.";  
 CC Science 269:496-512(1995).  
 CC -1- FUNCTION: REQUIRED FOR UTILIZATION OF FREE HEME AT LOW  
 CC CONCENTRATIONS (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Outer membrane (by similarity).  
 CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
 CC -1- SIMILARITY: LOW, TO THE CORRESPONDING PROTEIN IN SEROTYPE B OF  
 CC H. INFLUENZAE.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; U32696; AAC21789.1; -.
CC TIGR; H10113; -.
CC InterPro; IPR000531; TonB_boxC.
CC Pfam; PF00593; TonB_boxC.1.
CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
CC PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
CC Outer membrane; Transport; TonB box; Signal; Receptor;
CC Complete proteome.
CC SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 744 HEME/HMOPEXIN UTILIZATION PROTEIN C.
SQ SEQUENCE 744 AA; 85043 MW; 74D94P72E4AEC31 CRC64;

Query Match 5.4%; Score 111.5; DB 1; Length 744;
Best Local Similarity 21.9%; Pred. No. 3.8;
Matches 93; Conservative 58; Mismatches 144; Indels 129; Gaps 25;

QY 17 VAYTQQLYQPNESLPT-VELEPVVITIDKSGMALANRITQMPTHTTKVIYEEOI-QEQT 74
DB 16 VGLQNVTAQNSSNSNRELLPIIVNTNDDSNKLPGR-----SVLKQRNIEQXQAD 67
QY 75 GSRQLADVMAQLIPSLGVSSGTTSNFGQTMHGRQVQPLNGVPLTGSRDISRLNS---- 130
DB 68 NAANLNI-----LPGVNMAGGFPG-QQTIN-----ING--MGDAEDVRVQDGTAKS 113
QY 131 -----INPNQVARIIVLSGATS-IVGSGATGGLINIVTKSLEBEEQFETRIG--- 176
DB 114 FEKYQCSIFIEPELLRRVTVKGNYSPOYGGNGFAGTKFETKDARDFLQENQKIGGFL 173
QY 177 VHGSKLSSEGIGVQGVQSVAGVSENGVL-----ARLDVDYRTTGA---PDANGK--- 225
DB 174 KYGNNSNNQKTYSTALVQNEQKIDLLFGSVRNAGDYKRPDPSKILFSKNNQKTGLI 233
QY 226 -----IAPE-----PAQTDKQDSKLS-----VNTNVDWQDLDKQIN 258
DB 234 KLNWQISPHELLTLSSVYGIHKWEPPAKRDILPKESLSDIMRYGTDIAW-----KKKL- 288
QY 259 LAETHYNDKQDTYADYGN-----RLAVLFG-EKPSLN-----AIKGL--SLSEQ 301
DB 289 ----VYRDQKRDENYTLKYNLPENNPWINLSTQFSYKTTQNDMPKREASSGLVGLGNQ 344
QY 302 PKTT-----KSTFNINYH-HDDLWG-----NTINTNAYRRE-----KGRFYP 338
DB 345 SWITYSDLTFDINNTSTFNIKTTVHELLFGLQWLKNTRLNTMLYDKSKVRKADYNYGYFQ 404
QY 339 FVAP 342
DB 405 YNYP 408

RESULT 33
ID_FHUE_ECOLI STANDARD; PRT; 729 AA.
AC P16869; P77292;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE FhuE receptor precursor (Outer-membrane receptor for Fe(III)-coprogen,
DE Fe(III)-ferrioxamine B and Fe(III)-rhodotorulic acid).
GN FHUE OR B1102.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=K12;

```

```

RX MEDLINE=90286919; PubMed=2162465;
RA Sauer U., Hantke K., Braun V.;
RT "Sequence of the fhuE outer-membrane receptor gene of Escherichia
RT coli K12 and properties of mutants.";
RL Mol. Microbiol. 4:427-437(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [4]
RP SEQUENCE OF 37-50.
RC STRAIN=K12;
RX MEDLINE=87194585; PubMed=3032906;
RA Sauer M., Hantke K., Braun V.;
RT "Ferric-coprogen receptor FhuE of Escherichia coli: processing and
RT sequence common to all TonB-dependent outer membrane receptor
RT proteins.";
RL J. Bacteriol. 169:2044-2049(1987).
CC -|- FUNCTION: REQUIRED FOR THE UPTAKE OF IRON(III) VIA COPROGEN,
CC FERROXAMINE B, AND RHODOTORULIC ACID.
CC -|- SUBCELLULAR LOCATION: Outer membrane.
CC -|- INDUCTION: FOR INDUCTION THE TONB AND THE EXBB PROTEIN HAVE TO
CC BE ACTIVE.
CC -|- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
CC EMBL; X17615; CAA35616.1; -.
CC EMBL; AE000210; AAC74186.1; -.
CC EMBL; D90745; BAA35909.1; -.
CC EMBL; D90746; BAA35917.1; -.
CC PIR; S09262; S09262.
CC PIR; A26875; A26875.
CC EcoGene; EG10306; fhuE.
CC InterPro; IPR000531; TonB_boxC.
CC Pfam; PF00593; TonB_boxC; 1.
CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
CC PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
CC Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
CC Complete proteome.
FT SIGNAL 1 36
FT CHAIN 37 729 FHUE RECEPTOR.
FT SITE 42 49 TONB_BOX.
FT SITE 712 729 TONB_C-TERMINAL_BOX.
FT MUTAGEN 44 44 V->P: ABOLISHES TRANSPORT ACTIVITY.
FT MUTAGEN 46 46 V->P: ABOLISHES TRANSPORT ACTIVITY.
FT MUTAGEN 163 163 G->A,V,N,D,H: IMPAIRED TRANSPORT AND
FT GROWTH.

```



FT	CONFLICT	363	363	G -> C (IN REF. 1).
SO	SEQUENCE	729 AA;	81232 MM;	68ACBCE7D10F6CD CRC64;
	Query Match	5.4%;	Score 110;	DB 1; Length 729;
	Best Local Similarity	20.9%;	Pred. No. 4.6;	
	Matches	84;	Conservative	57; Mismatches 133; Indels 128; Gaps 22.
Oy	41	ITIDSQMALNRITOMPTHTVIVYBEQIQEQAOTSGROLADVMAGLIPSLGYSSGTTSNF	100	
Dd	62	VYTSAGTRKMTORTDIPDSVTIVSQOQRMEDQOL--QTIGEVNEN---TLGISKSQASD	116	
Oy	101	GQTMGRGVQVQ---PLINGPV-LTGSRDLSRQLNSINPQVARIETVLGATSIY-GSGAT	154	
Dd	117	RALYSRGQIDNNWVDGIPTYEPESHMNGDALSDM--ALPERVEYVRATGLMTGTGNP	174	
Oy	155	GGILINIYT-----KSDLLEE---QPETRIGVHSGKLSEGI-----GYVGQS-	194	
Dd	175	SAAIMWRKHATSREPKGVSAEYGSWMNERVYADLOSPLTEDEKIRARIVGGYQNDSM	234	
Oy	195	-----VAGVSE-----NCNVALARLDVDTYTTGG-----AFDANG	223	
Dd	235	LDRYNSEKTFPGSIGIYADADGLDILLTSAGAYEYORIDVNSPTWGGLPRMWTGDSNSYD-RA	293	
Oy	224	KRIAPPAQTDKODSKSLSVNTTNVMQDLDDKNINLATIH-----VNDKOD	269	
Dd	294	RSTAPDMAINDKEINK---VFMTLKQQPADTWQATLNATHSEVEFPSSRMYYDAYANKKD	350	
Oy	270	-----TDYAE--DY-----GNRLA-----VLFGEKPSINALIKGLSLSEQ	301	
Dd	351	GMLVGPYSNYGFGFDVVGSTGNNSGRKTKDALLDFADSGYEELFGRH--NLMGSGSYXNQ	408	
Oy	302	PRTYSTENINYHHDDLWCNTNTNATYAREBKRPYPFAVP	343	
Dd	409	-----NNRYESSMAN-----IFPDIGSFFNFNGNF	434	
	RESULT 34			
SP3	HUMAN	STANDARD;	PRT;	711 AA.
ID	SP3_HUMAN			
AC	Q02447;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Transcription factor Sp3 (SPR-2) (Fragment).			
GN	Sp3.			
OC	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euteria; Primates; Catarrhini; Hominiidae; Homo.			
OX	NCBI_TaxId=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND REVISIONS.			
RC	TISSUE=T-cell;			
RA	Kingsley C.; Winoc A.;			
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.			
	[2]			
RP	SEQUENCE OF 59-711 FROM N.A.			
RX	MEDLINE=93024366; PubMed=1341900;			
RA	Kingsley C.; Winoc A.;			
RT	"Cloning of GT box-binding proteins: a novel Spi multigene family			
RL	regulating T-cell receptor gene expression.";			
	Mol. Cell. Biol. 12:4251-4261 (1992).			
	[3]			
RP	SEQUENCE OF 15-711 FROM N.A.			
RC	TISSUE=Uterus;			
RX	MEDLINE=93087156; PubMed=1454515;			
RA	Hagen C.; Mueller S.; Beato M.; Sueke G.;			
RT	"Cloning by recognition site screening of two novel GT box binding			
RL	proteins: a family of Spi related genes.";			
	Nucleic Acids Res. 20:5519-5525 (1992).			
	[4]			
RP	FUNCTION			
RX	MEDLINE=97426517; PubMed=9278495;			
RA	Ihn H.; Tiojanowska M.;			

```

RT      "Sg3 is a transcriptional activator of the human alpha2(I) collagen
RL      gene.";
CC      Nucleic Acids Res. 25:3712-3717(1997).
CC      -1- FUNCTION: BINDS TO GT AND GC BOXES. PROMOTERS ELEMENTS. PROBABLE
CC      TRANSCRIPTIONAL ACTIVATOR.
CC      -1- SUBCELLULAR LOCATION: Nucleat.
CC      -1- TISSUE SPECIFICITY: UNIQUTOUSLY EXPRESSED.
CC      -1- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER
CC      PROTEINS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on iteas
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by, and for commercial
CC      entities requires a license agreement (See http://www.1db-sib.ch/announce)
CC      or send an email to license@1db-sib.ch.
CC      -----
DR      EMBL; M97191; AAA36630.2; -.
DR      EMBL; X68560; CAA48562.1; -.
DR      PIR; S26639; S26639.
DR      HSSP; P08047; 1SP1.
DR      TRANSFAC; T02338; -.
DR      Genew; HGNC:11208; SP3.
DR      MIM; 601804; -.
DR      InterPro; IPR000822; Znf_C2H2.
DR      Pfam; PF00096; Zf-C2H2; 3.
DR      ProDom; PD000003; Znf_C2H2; 1.
DR      SMART; SM00355; Znf_C2H2; 3.
DR      PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR      PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
KM      Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW      DNA-binding; Nuclear protein; Repeat.
FT      NON_TER 1
FT      DOMAIN 1 551 633 ZINC FINGERS.
FT      ZN_FING 551 575 C2H2-TYPE.
FT      ZN_FING 581 605 C2H2-TYPE.
FT      ZN_FING 611 633 C2H2-TYPE.
FT      CONFLICT 94 94 T -> A (IN REF. 3).
FT      CONFLICT 669 669 N -> K (IN REF. 3).
SQ      SEQUENCE 711 AA; 75102 MW; D1719FAC9B05A217 CRC64;

Query Match 5.3%; Score 108; DB 1; Length 711;
Best Local Similarity 20.9%; Pred. No. 6.1; Indels 108; Gaps 15
Matches 74; Conservative 50; Mismatches 122;

QY 52 NRITQMPHTKTYIEE-----OIQEQATGSRQ-----LADVMAQL-----PSIGVSSGTS 98
DB 37 NMWEVLSATPTTTIKDBAGNLVQIPSAIISGGYVLPQLNQLQOIFSVAPGSDSSNGTVS 96
QY 99 NFG-----QTMHGRQVQFLINGVPLTGSRD---ISRLNSI---MNVQV----- 136
DB 97 SVQYQVVIPIQISADQOYQ-----IGFGSSDNGINGNSSOIQIIPGSNQLTLASGTPS 151
QY 137 ARIEVL---SGATSIYSGATG---GLINITYKSDLDEEQTETRIGVHGSKLSSSEGTG 189
DB 152 ANIQNLILPQTGQVQVQGVVAIGSSSEFGQYQVAVNPVLGIAPGNITFPINISVLDLSIGLSG 211
QY 190 QVQGSVAVGSEKGNVLARLDVYRTTGAAGFADN-----GKRIAPAPATQKODSKSLV 243
DB 212 SSQYMTAGINAGHLL-----NTGQAMDSNDNSERTGENVSPDINET----- 253
QY 244 NTNVDVMDLDDKONINLALTHYNDKODPTDYAPRYGNRLAVLFGEKPSLNAIKGISLEQPK 303
DB 254 NIDTDLFPVTTSSSQQLPVJ-----IDNGILQQT 283
QY 304 TTKSTFNINVHHDDLWQNTINTNAYYRREKGFYPFVADPFSIAKALPIIQSKML 357
DB 284 NSLITTSIGGVHSSDLQGNVIGQSPVESETQ-----AQNIQVSTAQPVQHLQL 330

RESULT 35
C9DA_BACTP

```

```

ID C9DA_BACTP STANDARD; PRT; 1169 AA.
AC O06014;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Pesticidal crystal protein cry9Da (Insecticidal delta-endotoxin
DE CryIXD(a)) (Crystalline entomocidal protoxin) (132 kDa crystal
DE protein).
GN CRY9DA OR CRYIXD(A).
OS Bacillus thuringiensis (subsp. japonensis).
OS Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
OX NCBI_TaxID=129936;
RN SEQUENCE FROM N.A.
RP STRAIN=N141;
RA Asano S.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -|- FUNCTION: PROMOTES COLLOIDSMOTIC LYSIS BY BINDING TO THE MIDGUT
CC EPIITHELIAL CELLS OF INSECTS.
CC -|- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING
CC SPOREULATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART
CC OF THE SPORE COAT.
CC -|- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE
CC N-TERMINUS.
CC -|- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D85560; BAA19948.1; -
DR HSP; P07130; LDLC.
DR InterPro; IPR001178; Endotoxin.
DR Pfam; PF00555; endotoxin; 1.
KW Toxin; Sporulation.
SQ SEQUENCE 1169 AA; 132228 MW; 659AB257229DE5E9 CRC64;

Query Match 5.3%; Score 108; DB 1; Length 1169;
Best Local Similarity 22.2%; Pred. No. 12;
Matches 88; Conservative 49; Mismatches 119; Indels 140; Gaps 23;

QY 87 IPSLGVSSGTTSTNFGQTM-HGRQVF--LLNGVPLTGS-----RDISRLNSINP 133
DB 470 LPSLNEGDIIRNYSRHLSHITQYRFQATQSGSPSTVSANLPTCVWTHRDVLD-NTITA 528
QY 134 NOVARI-----EVLGATSIYSGATG-----LINIV 161
DB 529 NOITQLPLVKAVELSSGATVWKGFGTGGDVIRRTNTGGFGAIRVSVTGPLTORVIRFR 588
QY 162 TKSDLSEEFETRIGV-----HCKSLSSGIGVQ-----VQ 193
DB 589 YASTIDDFVTRKGGTINNFRTRTNWNGQESRYE--SYRTVEFTTPTNFTQSDIIRT 646
QY 194 SVAGVSENGV--LARIIDVYRTTGGAFDANGKRIAPAPATQDKDSK-----SLSV 243
DB 647 SIQGLSGNEVYLDRIE-----IPVNPAREADELEAKAKARQLFTTRDGLQV 698
QY 244 NTNVDHQLDDKQINIALTHYNDKQDTPADYGNBLAVLFGKPSLNAIKGLSLSEQPK 303
DB 699 NV-TDYQVDAQANLVSL---SDEQ-----YGHDKKMLL---EAVRAAKRLS-RERNL 743
QY 304 TTKSTEN-INYHDDLA---GNTINTNAYRREKGRFPYFFVAPFSIAKALPILQSM-NL 357
DB 744 LODPDFTNTSTENGWKAANGVTISEG-----GPFPGKRALQALASARENY 789
QY 358 P-----SATLIDAYTKAPQAPAYGVLOSSEKAEV 385
DB 790 PTYIYQKVASVLKPYTR---YELDGFVKSSQDLEI 822

```

```

RESULT 36
HGBB_HAEDU STANDARD; PRT; 972 AA.
AC Q47957;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemoglobin and hemoglobin-haptoglobin binding protein precursor
DE (Hemoglobin utilization protein A).
GN HGBB OR HUPA.
OS Haemophilus ducreyi.
OS Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OX Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=35000;
RX MEDLINE=96201585; PubMed=8613384;
RA Stevens M.K., Porcella S.F., Kleeney-Tait J., Lumley S., Thomas S.E.,
RA Norgard M.V., Radolf J.R., Hansen E.J.;
RT "A hemoglobin-binding outer membrane protein is involved in virulence
RT expression by Haemophilus ducreyi in an animal model.";
RL Infect. Immun. 64:1724-1735(1996).
CC -|- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX OF THE HOST AND IS REQUIRED FOR
CC HEME UPTAKE. MAY BE INVOLVED IN VIRULENCE.
CC -|- SUBCELLULAR LOCATION: Outer membrane.
CC -|- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U34048; AAC44054.1; -
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS00430; TONB-DEPENDENT REC_1; FALSE NEG.
DR PROSITE; PS01156; TONB-DEPENDENT REC_2; FALSE NEG.
KW Outer membrane; Transport; TonB box; Multigene family; Signal;
KW Receptor.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 972 HEMOGLOBIN AND HEMOGLOBIN-HAPTOGLOBIN
FT SITE 32 39 BINDING PROTEIN.
FT SITE 955 972 TONB C-TERMINAL BOX.
SQ SEQUENCE 972 AA; 110963 MW; E35BAE907A5F5349 CRC64;

Query Match 5.3%; Score 107.5; DB 1; Length 972;
Best Local Similarity 19.7%; Pred. No. 10;
Matches 68; Conservative 60; Mismatches 149; Indels 69; Gaps 17;

QY 10 LSLPLLSVAVTQOLYAQPNESLPTVELPWTITDKSGMALANRITQMPHTTKVIYEQI 69
DB 9 IYLCILGVANT--VVAEN--MQTEKLETVVSSDDSDVHNKN--VGEIKKNNAKSKQGV 63
QY 70 QEQATGSRQLADVMAQLIPSLGVSSGTTSTNFGQTMHGG---RQVQFLANGVP----- 117
DB 64 QDSRDLRVYETGVTV-----VEKGRFGSSGVAIRGVDENRVAVVVDGLHQAEITSSG 116
QY 118 ----LITGRDTRQLNSINPNQVARIIVLSGATSI-YSGGATGGLINIVTKSDLE----- 167
DB 117 FKLEPEGVGNFNTRNGVEVENLKQAVIQKGADAIKRTGSGISLGTVPSEFKDARDYLIDK 176
QY 168 ERQFETRIQ---VHGSKLSSEGIGVQVGSVAG-----VTAAGRYSDFDLLAVHTQHGNELR-- 213
DB 177 NYHFVYKGYSSADNOKLHS-----VTAAGRYSDFDLLAVHTQHGNELR--NYGYR 226

```

QY 214 TTGCAFDANGKRIAPPAQTDKODSKSLSVNTNVDMOLDKONINIALTHVNDKODDTYA 273  
 Db 227 HYDGSV-VKREKADPYKITKOSSL-----IKIGYOLNDTRFTLG---YDSHRTSRG 277  
 QY 274 PDYGRVLAFLGEEKPSLNAIKGLSLSEQPKTKSPFNINHHDDM 319  
 Db 278 TDWNAFTSYNG-GPFLKDVHRHTDQSNRKNISFYVE-NFTDNDPM 321

RESULT 37  
 TABP\_HUMAN STANDARD; PRT: 1729 AA.  
 AC Q9C0C2;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE 182 kDa tankyrase 1-binding protein.  
 GN TAB182 OR KIAA1741.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Placenta, and Testis;  
 RX MEDLINE=21950796; PubMed=11854288;  
 RA Selmiva H., Smith S.;  
 RT "The telomeric poly(ADP-ribose) polymerase, tankyrase 1, contains  
 RT multiple binding sites for telomeric repeat binding factor 1 (TRF1)  
 RT and a novel acceptor, 182-kDa tankyrase-binding protein (TAB182).";  
 RL J. Biol. Chem. 277:14116-14126(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=21082932; PubMed=11214970;  
 RA Nagase T., Kitano R., Hattori A., Kondo Y., Okumura K., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XIX.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RT for large proteins in vitro.";  
 RL DNA Res. 7:347-355(2000).  
 RN [3]  
 RP SEQUENCE OF 495-1729 FROM N.A.  
 RC TISSUE=Spleen;  
 RA Jikuya H., Takano J., Nomura N., Kitano R., Nagase T., Ohara O.;  
 RT "The nucleotide sequence of a long cDNA clone isolated from human  
 RT spleen.";  
 RL Submitted (JAN-2002) to the EMBL/Genbank/DBJ databases.  
 CC -1- SUBUNIT: Binds to the ANK repeat domain of TNKS1 and TNKS2.  
 CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic. Colocalizes with  
 CC chromosomes during mitosis, and in the cytoplasm with cortical  
 CC actin.  
 CC -1- TISSUE SPECIFICITY: Detected in testis, ovary, lung, skeletal  
 CC muscle, heart, prostate and pancreas, and at very low levels in  
 CC brain and peripheral blood leukocytes.  
 CC -1- PTM: ADP-ribosylated by TNKS1 (in vitro).  
 CC -1- CAUTION: Ref.3 sequence differs from that shown due to frameshifts  
 CC in position 1071, 1097 and 1467.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL, AF441771; AAM15531.1; -  
 CC EMBL, AB051528; BAB21832.2; ALT INT.  
 CC EMBL, AK074113; BAB64939.1; ALT\_FRAME.  
 CC MIM, 607104; -  
 CC Nuclear protein; Chromosomal protein; ADP-ribosylation.  
 KW DOMAIN 2 103 ARG/GLU/LYS/PRO-RICH (CHARGED).  
 FT DOMAIN 127 767 PRO-RICH.

FT DOMAIN 210 1572 ACIDIC.  
 FT DOMAIN 1010 1340 GLY-RICH.  
 FT DOMAIN 1450 1542 TANKYRASE-BINDING.  
 FT DOMAIN 1572 1729 ARG/GLU/LYS-RICH (CHARGED).  
 FT DOMAIN 1629 1635 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT DOMAIN 1723 1729 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
 FT CONFLICT 84 84 P -> L (IN REF. 2).  
 FT CONFLICT 322 322 S -> T (IN REF. 2).  
 FT CONFLICT 388 388 S -> P (IN REF. 2).  
 FT CONFLICT 554 554 O -> H (IN REF. 3).  
 FT CONFLICT 604 604 P -> S (IN REF. 1).  
 FT CONFLICT 1450 1450 F -> S (IN REF. 2).  
 SQ SEQUENCE 1729 AA; 181814 MW; C65F38FA37045C4A CRC64;

Query Match 5.3%; Score 107.5; DB 1; Length 1729;  
 Best Local Similarity 20.2%; Pred. No. 23;  
 Matches 81; Conservative 45; Mismatches 116; Indels 159; Gaps 15;

QY 16 SVAVTQ-----QLVAQNESLPTVELEPVV-----ITDKSGMALNR 53  
 Db 549 SMLTQKGDGDSQPQFPVAVLEPLTTEGTPLPQQAERYESQEPLAGQSPPLATR 608  
 QY 54 ITQMPHTKVIYEQ-----IQEQA----- 73  
 Db 609 EALPLLEPVLQEQPPAAPDQCVLFADAPBEGALPVEEAVTLARAEQTQARTQADL 668  
 QY 74 -----TGSRLADVMAQLPSLG-----VSSGTTSN--FGQTMHGR 107  
 Db 669 CQASPEPPSSSSSWLDDLLASPPSGGARGAGAEIKQTQSTSGELGMSQDL 728  
 QY 108 QVQPLANGVPLTGSRDISRQSLNPNQ-----VARIEVLSCA-----TSIYG 150  
 Db 729 QSEFGITGDP-----QPSSFPSMWCGASQDYGAGSPRGDPGLGRDWTSKYG 779  
 QY 151 SGA-TGGLINITYKSDLBEOPEFTRIGVHGKLSSEGGIYGQGVAGVSENGVLAAD 209  
 Db 780 QQAGGSGSTREWASRCIGIQEWEASSDOSVSAAG-----VLTAQD 822  
 QY 210 VYRTTGAFDANGKRIAPPAQTDKODSKSLSVNTNVDMOLDKONINIALTHVNDKOD 269  
 Db 823 -----RVGKPAQLGTQRSQEADVQ---DWEFRKDSQCTYSRDAEQLD 864

QY 270 TDYAP-----DYGNRLAVL-----FGKPSLNAIKGLSLSEQ 301  
 Db 865 QEGKGRDLSIGTSSRDVSLGDMFGRKDSLGAYASQADNEQ 905

RESULT 38  
 DNLJ MYCLE STANDARD; PRT: 694 AA.  
 AC 033102;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA ligase (EC 6.5.1.2) (Polydeoxyribonucleotide synthase [NAD+]).  
 GN LIG OR LIG OR M1705 OR MCB637.10.  
 OS Mycobacterium leprae.  
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1769;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TN;  
 RX MEDLINE=21128732; PubMed=11234002;  
 RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,  
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,  
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,  
 RA Holtroyd S., Hornby T., Jagers K., Lacroix C., Maclean J., Moule S.,  
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,  
 RA Rutter S., Seeger K., Simon S., Skelton J., Skelton J., Squares R.,  
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,  
 RA Barrell B.G.;

RT "Massive gene decay in the leprosy bacillus.";  
 RL Nature 409:1007-1011(2001).  
 CC -!- LINKAGES: THIS PROTEIN CATALYZES THE FORMATION OF PHOSPHODIESTER  
 CC STRANDED DNA USING NAD AS A COENZYME AND AS THE ENERGY SOURCE FOR  
 CC THE REACTION. IT IS ESSENTIAL FOR DNA REPLICATION AND REPAIR OF  
 CC DAMAGED DNA (BY SIMILARITY).  
 CC -!- CATALYTIC ACTIVITY: NAD(+) + (deoxyribonucleotide)(N) +  
 CC {deoxyribonucleotide}(M) = AMP + nicotinamide nucleotide +  
 CC {deoxyribonucleotide}(N+M).  
 CC -!- SIMILARITY: BELONGS TO THE NAD-DEPENDENT DNA LIGASE FAMILY.  
 CC -!- SIMILARITY: CONTAINS 1 BRCT DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z99263; CAB16425.1; -;  
 CC EMBL; AL583923; CAC30658.1; -;  
 CC HSP; 087703; LB04.  
 CC Leproma; ML1705; -;  
 CC InterPro; IPR001357; BRCT.  
 CC InterPro; IPR004150; DNA\_ligase\_OB.  
 CC InterPro; IPR001679; DNAligase.  
 CC InterPro; IPR003583; HH1.  
 CC InterPro; IPR000445; HhH.  
 CC InterPro; IPR004149; Znf\_DNAligase\_C4.  
 CC Pfam; PF00533; BRCT; 1.  
 CC Pfam; PF00633; HH1; 1.  
 CC Pfam; PF01653; DNA\_ligase\_N; 1.  
 CC Pfam; PF03119; DNA\_ligase\_ZBD; 1.  
 CC Pfam; PF03120; DNA\_ligase\_OB; 1.  
 CC ProDom; PD003944; DNAligase; 1.  
 CC SMART; SM00292; BRCT; 1.  
 CC SMART; SM00278; HH1; 1.  
 CC SMART; SM00532; LIGANC; 1.  
 CC TIGRfam; TIGR00575; dn1j; 1.  
 CC PROSITE; PS0172; BRCT; 1.  
 CC PROSITE; PS01055; DNA\_LIGASE\_N; 1.  
 CC PROSITE; PS01056; DNA\_LIGASE\_Z; FALSE NEG.  
 CC Ligase; DNA repair; DNA replication; NAD; Complete proteome.  
 CC DOMAIN 613 677 BRCT.  
 CC BINDING 123 123 AMP (BY SIMILARITY).  
 CC SEQUENCE 694 AA; 75912 MW; CPE712B6AFDA9D5 CRC64;

Query Match 5.2%; Score 106; DB 1; Length 694;  
 Best Local Similarity 20.1%; Pred. No. 8.1;  
 Matches 67; Conservative 58; Mismatches 116; Indels 92; Gaps 17;  
 QY 35 ELEPVVITIDKSGMALANRI--TQPHHTTKVIEEQIQOATGSRQOLADVMAQLIPSLGV 92  
 DB 293 EDIGVVKWIDE--VTQRLGTSRAPRAVAIKYLPPEAQ--KLLDIRV----- 339  
 QY 93 SSGTTSNFGQTHGROVQF--LLNGVPLTGSROISRLNSINQVARIIEVLGATSIYGS 151  
 DB 340 -----NVGRT--GRVTPPAFTPVKVGSTVGQATLH--NPSEVKRGVLIGDTVIRK 389  
 QY 152 G-----ATGGLINIVTKSDLE-----BEQETRIG----- 176  
 DB 390 AGDVIPVLGVVLDLGDGSEBPMPTTCCGTTLAPEKEGDADIRCPNARSCPQOLRE 449  
 QY 177 -----VHGKLSSEGIGYQVGQSVAG--VSENGNV-----LARLDVDTYTTGAFDA 221  
 DB 450 RVFHVASRALDIQGLGYBAGVALLAAQVITSEGDIFTTEKALLTEL-FRNKAGELSA 508  
 QY 222 NGKRIAPEAQTDKQDSKLSLVNVDWQDDQKNINLALHYNDKQDQTDYAPDYNRLA 281  
 DB 509 NGKRL--VNVDKAKTAPL-----WRV-----LVALSIRHVGTPAARALATEFGSVA 554

QY 282 VLFGEKPSLNAIKGLSLSEQPTTKSTFNINTH 314  
 DB 555 ILAASPEQLAAVEGVGTITAAAVTE-WFTVDWH 586  
 RESULT 39  
 TBPI\_HAEIN STANDARD; PRT; 912 AA.  
 ID TBPI\_HAEIN AC P44970;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable transferrin-binding protein 1 precursor.  
 GN TBPA OR TBPI OR HI0994.  
 OS Haemophilus influenzae.  
 OS Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 OC NCBI\_TaxID=727;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=rd / KW20 / ATCC 51907;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.P.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-P., Dougherty B.A., Merrick J.M.,  
 RA McInerney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,  
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae  
 RT Rd.";  
 RL Science 269:496-512(1995).  
 CC -!- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR  
 CC TRANSFERRIN UTILIZATION (BY SIMILARITY).  
 CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).  
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U32780; AAC22656.1; -;  
 CC TIGR; HI0994; -;  
 CC InterPro; IPR000531; TonB\_boxC.  
 CC Pfam; PF00593; TonB\_boxC; 1.  
 CC PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; 1.  
 CC PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
 CC Outer membrane; Receptor; Signal; TonB box; Complete proteome.  
 CC SIGNAL 1 23 POTENTIAL.  
 CC CHAIN 24 912 PROBABLE TRANSFERRIN-BINDING PROTEIN 1.  
 CC SITE 50 57 TONB BOX.  
 CC SITE 895 912 TONB C-TERMINAL BOX.  
 CC SEQUENCE 912 AA; 103233 MW; 294C08991AC52C2 CRC64;  
 Query Match 5.2%; Score 106; DB 1; Length 912;  
 Best Local Similarity 24.2%; Pred. No. 12;  
 Matches 48; Conservative 36; Mismatches 88; Indels 26; Gaps 8;  
 QY 18 AVTQOLYAPQNESLPTVEPVITIDKSGMALANRIITQPHHTTKVIEEQIQOATGSR 77  
 DB 34 AISSEVDQTQSTED-----SELETISVTAERKVRKORKEVTLGLKLIKT-SESTISREQVLNIR 89  
 QY 78 QLADVMAQLIPSLG-VSSGTTSNFGQTHG---ROVQFLNGVPLTGSRDISRQL----- 128  
 DB 90 DL-----TRYDPCISVVEQGRGASSGYSIRGMDRNRVALLVDGLPQTQSVVQSPVARS 145  
 QY 129 -----NSINPNQVARIIEV-LSGATSIYGSATGGLINIVTKSDLEBEQETRIGVH-G 179

Db 146 YSGTAINIEYENKAVAISSKGGSSSEYGNALGAVTFOSKSNADILEGKSWGIQTK 205  
 QY 180 SKLSEGIYGVGVGAVG 197  
 Db 206 NAYSSKNGKFTSLAVAG 223

RESULT 40  
 LBPA\_NEIMB  
 ID LBPA\_NEIMB STANDARD; PRT; 943 AA.  
 AC 006379; 090YKS; (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Lactoferrin binding protein A precursor (Iron-regulated outer membrane protein A).  
 GN LBPA OR IRGA OR NMB1540.  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxId=491;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BNV / Serogroup B;  
 RX MEDLINE=94011384; PubMed=8406871;  
 RA Petersen A., van der Ley P., Poolman J.T., Tommassen J.;  
 RT "Molecular characterization of the 98-kilodalton iron-regulated outer membrane protein of Neisseria meningitidis.";  
 RL Infect. Immun. 61:4724-4733(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MCS8 / Serogroup B;  
 RX MEDLINE=20175755; PubMed=10710307;  
 RA Tetteijn H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.R., Riesen J.A., Ketchum K.A., Hood D.W., Breden J.F., Dodson R.J., Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Ciecko A., Parksey D.S., Blair E., Cifton H., Clark E.B., Cotton M.D., Ulteback T.R., Khouli H., Qin H., Vamathevan J., Gill J., Scariato V., Maignani V., Pizza M., Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;  
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain MCS8.";  
 RL Science 287:1809-1815 (2000).  
 CC -1- FUNCTION: UNKNOWN. MAY BE AN IRON-SIDEROPHORE RECEPTOR.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane.  
 CC -1- INDUCTION: BY IRON STARVATION.  
 CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL; X69214; CAA49148.1; -  
 DR EMBL; AE002504; AAF41895.1; -  
 DR TIGR; NMB1540; -  
 DR InterPro; IPR000531; TONB\_boxC.  
 DR Pfam; PF00593; TONB\_boxC.1.  
 DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; FALSE\_NEG.  
 DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
 KW Outer membrane; Iron transport; Transport; Tonb box; Signal; Receptor; Complete proteome.  
 KM  
 FT SIGNAL 1 27  
 FT 943 943 LACTOFERRIN BINDING PROTEIN A.  
 FT SITE 876 943 TONB C-TERMINAL BOX.  
 FT 8 P -> Q (IN REF. 1).  
 FT CONFLICT 16 18 IAT -> VAA (IN REF. 1).  
 FT CONFLICT 22 22 A -> S (IN REF. 1).  
 FT CONFLICT 26 31 QAGAT -> NPTAA (IN REF. 1).

FT CONFLICT 43 43 I -> V (IN REF. 1).  
 FT CONFLICT 64 64 V -> A (IN REF. 1).  
 FT CONFLICT 233 233 R -> H (IN REF. 1).  
 FT CONFLICT 243 243 E -> A (IN REF. 1).  
 FT CONFLICT 247 247 D -> N (IN REF. 1).  
 FT CONFLICT 257 269 DIKRTREPFPSV -> GIKPSSEGEYFLA (IN REF. 1).  
 FT CONFLICT 273 275 RES -> SEL (IN REF. 1).  
 FT CONFLICT 281 281 L -> V (IN REF. 1).  
 FT CONFLICT 284 286 YGK -> NGN (IN REF. 1).  
 FT CONFLICT 313 313 Q -> M (IN REF. 1).  
 FT CONFLICT 389 389 E -> K (IN REF. 1).  
 FT CONFLICT 420 423 KNLV -> QXLI (IN REF. 1).  
 FT CONFLICT 455 455 A -> K (IN REF. 1).  
 FT CONFLICT 546 546 K -> N (IN REF. 1).  
 FT CONFLICT 564 572 STGDENNQ -> YSDYDKG (IN REF. 1).  
 FT CONFLICT 658 658 L -> V (IN REF. 1).  
 FT CONFLICT 667 667 V -> L (IN REF. 1).  
 SQ SEQUENCE 943 AA; 105680 MW; BD569BCACFC01A84 CRC64;

Query Match 5.2%; Score 106; DB 1; Length 943;  
 Best Local Similarity 19.5%; Pred. No. 12; Mismatches 156; Indels 166; Gaps 21;  
 Matches 93; Conservative 63;

QY 1 MEHSHYFQWLSPLLSVAVTQOLY-AQPNESLP-----TVELPVIYITDKSGMAL--ANR 53  
 Db 1 MKKGIGFP-LTLTALAIATAPAYAAQAGAPDAAQGTSLKEITVPAKVGRRSEATG 59  
 QY 54 ITQMPHTTVIYEEQIQEQATSSROLADYMAQLISLG-VSSGTTSNFGQTHG--ROY 109  
 Db 60 LKRIYVTSSTLNKEQV---LGIRDL---TRYDGVAAVEQNGASGVYSIRGYDKRV 111  
 QY 110 QPLNGVPLT-----GSRDISRQNSINPNQVARIETVLSGA-TSITYSGATG 156  
 Db 112 AVSVCVAQIQAFVYQGSLSYGSGGSGALIEYENISTYETIDKAGSSDHGSGALG 171  
 QY 157 LNIYTK--SDU-----EEQFETRIGVHSGKLSSEGIYGVGVGAVG 197  
 Db 172 AVAFRKEAADISDGKSGIQAKTAYGSKNPFMSLGAGSSKQMBELLIRTERQGE 231  
 QY 198 VESNGV---LARI-----DVDYRTTGAFDANGKR----- 225  
 Db 232 TRPHGIADGVYGIIDRDAFRQYDIKRTREPFPSVSEGESKPVAKIAGYKTLNQ 291  
 QY 226-----IAPPAQT-----DKDSSKLSYNTNV 247  
 Db 292 LIRWYKERIEQNQPLSAEEAQVREARQAHENLSAQATYGGRIILPDPYRSGLAKL 351  
 QY 248 DWQL-----DDKONIMALTNYNDKQDTYAPDYGNRLAVLFG 286  
 Db 352 GYRFGGRHYGVGFEDTKQRYDIRMTEKQYVGTDBAEKFRKSGVYDDDFRDGLYFVP 411  
 QY 287 ----XPSINAIGKLSLSQPKTKSTFNINYHDDLWGNITNTNAYYRREKGRPYF 339  
 Db 412 NIEEWGDKNLVIRGIGL---KYSRTKF-IDEBH-----RRRRMGLLYRY 451

RESULT 41  
 POPA\_PSEPU  
 ID POPA\_PSEPU STANDARD; PRT; 819 AA.  
 AC P25184;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE Ferric-pseudodactin 358 receptor precursor.  
 GN POPA.  
 OS Pseudomonas putida.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 OC Pseudomonas.  
 OX NCBI\_TaxId=303;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MCS358;

RX MEDLINE=91260449; PubMed=1646376;  
 RA Bitter W., Marugg J.D., de Weger L.A., Tommassen J., Weisbeek P.J.;  
 RT "The ferric-pseudobactin receptor Pupa of Pseudomonas putida WCS358:  
 RT homology to TonB-dependent Escherichia coli receptors and specificity  
 RL of the protein.";  
 RL Mol. Microbiol. 5:647-655(1991).  
 CC -!- FUNCTION: SPECIFIC RECEPTOR FOR THE SIDEROPHORE FERRIC  
 CC PSEUDOBACTIN 358.  
 CC -!- SUBCELLULAR LOCATION: Outer membrane.  
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X56605; CAA39942.1; -.  
 DR PIR: S15169; S15169.  
 DR InterPro: IPR000531; TonB\_boxC.  
 DR Pfam: PF00593; TonB\_boxC; 1.  
 DR PROSITE: PS00430; TONB\_DEPENDENT\_REC\_1; 1.  
 DR PROSITE: PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
 KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor.  
 FT SIGNAL 1 47 POTENTIAL.  
 FT CHAIN 48 819 FERRIC-PSEUDOBACTIN 358 RECEPTOR.  
 FT SITE 115 122 TONB BOX.  
 FT SITE 802 819 TONB C-TERMINAL BOX.  
 SQ SEQUENCE 819 AA; 91015 MW; 1B36164F86207951 CRC64;

Query Match 5.2%; Score 105.5; DB 1; Length 819;  
 Best Local Similarity 20.2%; Pred. No. 11;  
 Matches 56; Conservative 50; Mismatches 102; Indels 69; Gaps 13;

QY 54 ITQMPHTTKVIEEQIQEQATGSRQADVMAQLIPSLGVSSGTTSTNFGQTMGRQVQ-FL 112  
 DB 167 IRETPTITVTRQMDQHLGS--MNEVLTO-TPGITSQDGERFNYSRGSAINIYQ 223  
 QY 113 LAGVPLTSGRDISRQLNS--INPNQARIEVLGSGSIY-GSGATGGLINIVTKSLEEE 169  
 DB 224 FDGV-ITYQDNQTRNPSLMDVGLYDRIEIVRGATGLATGADGPSAVNVIRKPTRE- 281  
 QY 170 QPETRIGVHGSKLSSEGIG-----YQGVQSVAG-VSENGVY----- 204  
 DB 282 -FKSHI-----QAGVGSWDYRAEADVSGPLTDDGRVGRFFAAKQDNHTFMDWYTQ 332  
 QY 205 -----LARDVDYRTTGGAPDANGKRIAPAPQTDKQDSKLSLVNTNV 247  
 DB 333 DRDLVGVVEADVDTTVARFGIDRGT-----YKVGAPGVPIIYNTGQPTNFSRSTSSDA 388  
 QY 248 DWQLDDKQINLALTHYNDKQDTPADPYGNRLAVLF 284  
 DB 389 RWGYDD-----YTTNTYTGLEQLAHDHOFKLAAY 420

RESULT 42  
 ID PBUA\_PSESP STANDARD; PRT; 826 AA.  
 AC Q08017;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ferric-pseudobactin M114 receptor pbua precursor.  
 GN PBUA.  
 OS Pseudomonas sp. (strain M114).  
 OC Bacteria; Proteobacteria.  
 OX NCBI\_TaxID=306;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94104606; PubMed=8277948;  
 RA Morris J., Donnelly D.F., O'Neill E., McConnell F., O'Gara F.;

RT "Nucleotide sequence analysis and potential environmental  
 RT distribution of a ferric pseudobactin receptor gene of Pseudomonas  
 RT sp. strain M114.";  
 RL Mol. Gen. Genet. 242:9-16(1994).  
 CC -!- FUNCTION: SPECIFIC RECEPTOR FOR THE SIDEROPHORE FERRIC  
 CC PSEUDOBACTIN M114.  
 CC -!- SUBCELLULAR LOCATION: Outer membrane.  
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X73412; CAA51812.1; -.  
 DR InterPro: IPR000531; TonB\_boxC.  
 DR Pfam: PF00593; TonB\_boxC; 1.  
 DR PROSITE: PS00430; TONB\_DEPENDENT\_REC\_1; 1.  
 DR PROSITE: PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
 KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor.  
 FT SIGNAL 1 44 BY SIMILARITY  
 FT CHAIN 45 826 FERRIC-PSEUDOBACTIN M114 RECEPTOR PBUA.  
 FT SITE 110 119 TONB BOX.  
 FT SITE 809 826 TONB C-TERMINAL BOX.  
 SQ SEQUENCE 826 AA; 90388 MW; 5D440D6AC797E4CB CRC64;

Query Match 5.2%; Score 105.5; DB 1; Length 826;  
 Best Local Similarity 21.6%; Pred. No. 11;  
 Matches 86; Conservative 59; Mismatches 144; Indels 109; Gaps 23;

QY 41 ITIDKSGMALNRITQMPHTTKVIEEQIQEQATGSRQADVMAQLIPSLGVSSGTTSNP 100  
 DB 152 VTIGK-----VHSLKETQSVTVTKMLDDQNLNT--IEQWKEK---TPGITVVDSPMG 202  
 QY 101 GQTMGR-----QVFLNGVPL-TGSRDISRQLNSINPNQVA--RIEVLGSGTSIY-GSG 152  
 DB 203 GKIFYSRGFRMSGQYQYDGVPLDIGSVY--QADSFNSDMAIYDRVLRGAAGMMKGAG 260  
 QY 153 ATGGLINIVTK--SLEEEQFETRIGVHGSKLSSEGIGVQGVQSVAG--VSENGVRLARL 208  
 DB 261 GTAGGVNFRKRGQDTAHTQLSLSGATWDN-----YR-GQVDTGGLNDSGTIRGRA 311  
 QY 209 DVDYRTTGGAFDANGK-----IAPE-----PAQTDKQ 236  
 DB 312 VTEQTRQYFYDVGSRKQIYYGALDFDLSPDTLGLGFAMEDVDATPCWGLPRYADGS 371  
 QY 237 D---SKLSLVNT-----NVDWLDDKQINLALTHYNDKQDTPADPYGNR 279  
 DB 372 DLHLKRSTCLANTAMNQRKATYFADLKHQPNDDWSLKVAGVYSRNTQDMEYAFPSG-- 429  
 QY 280 LAVLFGKPSLNAIKGLSLSEOPKTKSTFNINHYHD---DLWGN-----TINTNAYYRRE 332  
 DB 430 -AVPVG---ATATNTMLGSIYDYDQDYDFADVQDKFADFAGQOQHETIGANA-SRSH 483  
 QY 333 KGRFYFPFVAPPSIAKALPILOSMNLP SATDAYTKAPQ 370  
 DB 484 KDDFY-----AVALPQRQNV-----LDENHHIPQ 508

RESULT 43  
 ID HRPN\_ERWCH STANDARD; PRT; 340 AA.  
 AC Q47278;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Harpin (Harpin-ECH).  
 GN HRPN.  
 OS Erwinia chrysanthemi.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

```

OC Pectobacterium.
OX NCBI_TaxID=556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EC16 / AC1450;
RX MEDLINE=96172740; PubMed=8589405;
BAUER D.W., Wei Z.M., Beer S.V., Collier A.;
RT "Erwinia chrysanthemi harpinCh: an elicitor of the hypersensitive
RT response that contributes to soft-rot pathogenesis.";
RL Mol. Plant Microbe Interact. 8:484-491(1995).
CC -1- FUNCTION: ELICITS THE HYPERSENSITIVE RESPONSE (HR) IN THE PLANT
CC UPON INFECTION. HARPIN ELICITS HR IN NON-HOSTS AND IS ALSO
CC REQUIRED FOR PATHOGENICITY IN HOST PLANTS.
CC -1- SUBCELLULAR LOCATION: SECRETED; VIA THE HRP SECRETION PATHWAY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC DR EMBL; U39897; AAC31978.1; -.
CC HYPERSENSITIVE response.
CC FT DOMAIN 1 203 GLY-RICH.
CC SEQUENCE 340 AA; 34274 MW; E4D82C3731EF4C5 CRC64;
SQ
Query Match 5.1%; Score 104.5; DB 1; Length 340;
Best Local Similarity 21.3%; Pred. No. 3.8;
Matches 74; Conservative 42; Mismatches 113; Indels 119; Gaps 16;
QY 73 ATGSRQLADVMAQLISLGVSS---GTTSNFGQTM-HGRVQFLNGVPLTSGRDSR-- 126
DB 52 ALTSMFMFGALAQ---GLASSKGLGMSNOLGSPKNGAQASNLISVPSGSDALSKAF 108
QY 127 -----QLNSINPQVARIYELVSGATSIYSGATGGLINITYKSDLER 168
DB 109 DKALDILLGHDTVTKLINQNSQLANSLMANSQMTQGNMNAFGSGVNNALSSI----- 160
QY 169 EGFETRIGVHGSLSSEGGVGVGVSVAGSEKGNVLAARDVDYRTTGAFDANGKRIAP 228
DB 161 -----LGNGIGQSMGSPFSOPS--LGAGGLGGLGAGAFNOLGNAIGM 200
QY 229 EPAQDTKODSKSLSVNTNVDMDLDKQNTINLALTHYNDKODTGYAPDYG---NRLAVLFG 285
DB 201 GVGQ---NAALSALSNVSTHYDGNRR-----HFVDKEDRMALVEIGQFMQYPIFIQ 249
QY 286 EKPSLNAIKGLSLSEQPKTKTSTFNINHYHDDLGMGNTINTNAVYRREKGRFPVAFPSI 345
DB 250 -KREYVK-DGWS---SPKT-----DDKSM----- 268
QY 346 AKALPLQSMNLPSATLDATKAPQARAYGVLOSSEKAELYGVPPNLN 393
DB 269 AKALSKPDDDGMTGASMDKF-----ROAMGMI-----KSAVAGDTGNTN 307
GN
RESULT 44
OMPL PHOPR STANDARD; PRT; 341 AA.
ID OMPL PHOPR STANDARD; PRT; 341 AA.
AC Q52581;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Porin-like protein P precursor.
GN OMPL.
OS Photobacterium profundum (Photobacterium sp. (strain SS9)).
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae;
OC Photobacterium.
OX NCBI_TaxID=74109;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96326353; PubMed=8759872;

```

```

RA Welch T.J., Bartlett D.H.;
RT "Isolation and characterization of the structural gene for OmpL, a
RT pressure-regulated porin-like protein from the deep-sea bacterium
RT Photobacterium species strain SS9.";
RL J. Bacteriol. 178:5027-5031(1996).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane
CC (Potential).
CC -1- INDUCTION: IN RESPONSE TO ELEVATED HYDROSTATIC PRESSURE.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC DR EMBL; U59311; AAB50064.1; -.
CC DR InterPro; IPR003229; OMP_2.
CC DR InterPro; IPR001702; Porin_gam-ve.
CC DR Pfam; PF00267; Gram-ve_porins; 1.
CC KW Transmembrane; Porin; Signal; Outer membrane.
CC FT SIGNAL 1 21
CC FT CHAIN 22 341 PORIN-LIKE PROTEIN L.
CC SEQUENCE 341 AA; 36672 MW; ED3D66A93DEB585 CRC64;
SQ
Query Match 5.1%; Score 104.5; DB 1; Length 341;
Best Local Similarity 23.9%; Pred. No. 3.8;
Matches 57; Conservative 26; Mismatches 74; Indels 81; Gaps 12;
QY 158 INIVTKSDLEBEOFRIGVHGSKLS-----EGIGYGVGSGVAGV----- 198
DB 69 INVAGKTIDTEDFY--GVGFPEKPESSADSDNDETRVAYAVGSGVQGLVYGADGSLGM 126
QY 199 -----SEGNVLAARDVDYRTTG---GAPANGKRIAPAPAQDTKQDSKL 241
DB 127 LTFDFIDMAYHNGENKLAAD---RTDNMLSYVGSFPLNG-----DNL 168
QY 242 SVNTN-VDMQLDQKQNTINLALTHYNDKODTGYAPDYGNRLAVLFGKPSLNAIKGLSLSE 300
DB 169 TYKANYVFGSGSDENEGYSAAAM-----YAMDGLGFGAGYGEQDQSSKNG---N 215
QY 301 QPKTKSTFN-INHYHDDL-----GNTINTNAVYRREKGRFPVAFPSIAKAL 349
DB 216 EDKTGQANGALSTYTSIDFYSGLVQDSNRTVYNNDLIDESTG--YEFAAATYTGAV 271
GN
RESULT 45
FCUA YEREN STANDARD; PRT; 758 AA.
ID FCUA YEREN STANDARD; PRT; 758 AA.
AC Q05202;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Ferrichrome receptor fcua precursor.
GN FCUA.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=630;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-5 AND 37-46.
RC STRAIN=serotype O:8;
RX MEDLINE=93211281; PubMed=8384682;
RT Koehnlik R., Hanke K., Braun V.;
RT "The TonB-dependent ferrichrome receptor fcua of Yersinia
RT enterocolitica: evidence against a strict co-evolution of receptor
RT structure and substrate specificity.";
RL Mol. Microbiol. 7:183-193(1993).
CC -1- FUNCTION: RECEPTOR FOR THE HYDROXAMATE SIDEROPHORE, FERRICHROME.
CC BINDS ALSO TO MOST OTHER FERRICHROME DERIVATIVES EXCEPT ENNAVITO.
CC FERRICHROME AND FERRIC RHODOTORULATE.
CC -1- SUBCELLULAR LOCATION: Outer membrane.

```





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 25, 2002, 19:20:13 ; Search time 81 Seconds  
(without alignments)  
1017.517 Million cell updates/sec

Title: US-09-889-746-2

Perfect score: 2047  
Sequence: 1 MRHSHYFQWLSIPLLSVAVT.....SKAEVLGRVPLNKPKALP 400

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

SPTREMBL\_21: \*  
1: sp\_archaea: \*  
2: sp\_bacteria: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phase: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp\_vertebrate: \*  
14: sp\_unclassified: \*  
15: sp\_virus: \*  
16: sp\_bacteriap: \*  
17: sp\_archaeap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	527	25.7	813	16	Q913X9
2	503.5	24.6	742	16	Q9HVC0
3	497	24.3	812	16	Q8XSO1
4	493	24.1	725	2	Q934J5
5	491	24.0	726	16	Q8ZHB0
6	472	23.1	732	2	Q9XCH0
7	472	23.1	732	2	Q93PE2
8	453.5	22.2	851	16	P72594
9	453	22.1	867	16	Q8YZR0
10	441	21.5	705	2	Q9ZNM3
11	400.5	19.6	880	16	Q8YTY2
12	333.5	16.3	536	16	Q8YTY2
13	326.5	16.0	882	16	Q8YUX2
14	324.5	15.9	840	2	Q44003
15	205	10.0	646	2	P95494
16	204.5	10.0	657	16	Q8RHN9

17	202.5	9.9	687	2	Q9XD18	Q9xd18 bacteroides
18	200	9.8	726	16	Q8Z4H1	Q8z4h1 salmonella
19	199.5	9.7	725	2	Q9R019	Q9r019 escherichia
20	199.5	9.7	727	2	Q54507	Q54507 salmonella
21	196.5	9.6	621	16	Q8XWS0	Q8xws0 ralsstonia s
22	196.5	9.6	725	2	Q93K73	Q93k73 escherichia
23	195	9.5	726	2	Q8RMB4	Q8rmb4 salmonella
24	194	9.5	726	16	Q8ZMN0	Q8zmn0 salmonella
25	189.5	9.3	688	2	Q93SS7	Q93ss7 pseudomonas
26	187.5	9.2	698	16	Q8YRY2	Q8yry2 anabaena sp
27	187	9.1	867	16	Q8YTP0	Q8ytp0 anabaena sp
28	186	9.1	747	2	Q86903	Q86903 salmonella
29	182	8.9	872	16	Q8YTT7	Q8ytt7 anabaena sp
30	181.5	8.9	858	16	Q8YMK9	Q8ymk9 anabaena sp
31	180.5	8.8	708	2	Q9XBV1	Q9xbv1 porphyromon
32	180	8.8	853	16	Q8YU26	Q8yu26 anabaena sp
33	178.5	8.7	757	2	Q9KIB0	Q9kib0 porphyromon
34	178.5	8.7	802	16	Q8YV28	Q8yv28 anabaena sp
35	175.5	8.6	715	16	Q8RFD8	Q8rfd8 fusobacteri
36	173	8.5	732	2	Q86424	Q86424 pantoea agg
37	171	8.4	747	16	Q8X901	Q8x901 escherichia
38	170.5	8.3	1097	2	P72196	P72196 porphyromon
39	169.5	8.3	696	2	Q07651	Q07651 campylobact
40	168.5	8.2	696	16	Q9PE3	Q9pe3 campylobact
41	168.5	8.2	709	16	Q9PM61	Q9pm61 campylobact
42	168	8.2	820	16	Q8YTW7	Q8ytw7 anabaena sp
43	167.5	8.2	726	2	Q9X7K9	Q9x7k9 rhizobium l
44	167.5	8.2	747	2	Q93J93	Q93j93 rhizobium l
45	166.5	8.1	742	2	Q68598	Q68598 pseudomonas

## ALIGNMENTS

RESULT 1	Q913X9	PRELIMINARY;	PRT;	813 AA.
ID	Q913X9			
AC	Q913X9;			
DT	01-MAR-2001 (TREMBLrel. 16, Created)			
DT	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DT	01-OCT-2001 (TREMBLrel. 18, Last annotation update)			
DE	Probable siderophore receptor.			
GN	PAL165.			
OS	Pseudomonas aeruginosa.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;			
CC	Pseudomonas.			
OK	NCBI_taxid=287;			
RM	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 15692 / PA01;			
RX	MEDLINE=20437337; PubMed=10984043;			
RA	Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,			
RA	Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,			
RA	Garber R.L., Goltz L., Tolentino E., Westbrock-Wadman S., Yuan Y.,			
RA	Brody L.L., Coulter S.N., Folger K.R., Kas A., Laibis K., Lim R.M.,			
RA	Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,			
RA	Reiser J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,			
RT	"Complete genome sequence of Pseudomonas aeruginosa PA01, an			
RT	opportunistic pathogen."			
RT	Opportunistic pathogen."			
RL	Nature 406:959-964(2000).			
DR	EMBL; AB004565; AAC04754.1; -			
DR	InterPro: IPR000531; TonB_boxc.			
DR	Pfam: PF00593; TonB_boxc; 1.			
KW	Receptor; Complete proteome.			
SQ	SEQUENCE 813 AA; 89127 MW; 3861AB56E3105299 CRC64;			
Query Match	25.7%; Score 527; DB 16; Length 813;			
Best Local Similarity	34.2%; Pred. No. 2.6e-26;			
Matches 137; Conservative 68; Mismatches 144; Indels 52; Gaps 9;				
Qy	2 RSHSHYFQWLSIPLLSVAV..TQQLVAQPNESLPTVELEPVYITIDSGMALANRITOMPRT 60			
Db	94 RDEHTF--IVVDASEAAVDATQARSEP-----LDMEQWEITASRTSDIVSATRQ---- 141			

```

QY 61 TKVYIEQIQEQATGSRQLADVAQQLPSLGVSSGTTSNFGQTMHGRQVQVFLNGVPLTG 120
DB 142 STVIEHAQLEELRQGGDSLATLAKAVPGMSRSSRTITETGQTLRGLRSLVMVMDGVPLNT 201
QY 121 SDISRLQNSINPNOVARIEVLGATSIYSGATGGLINIVTKSDLEEFQFETRIGVHG- 179
DB 202 NRDSSRLNANIIPALIERIEVINGSAIYSGATGGIISITTPAGGENRAETRLSATSP 261
QY 180 -SKLSSEGIGYQVQSVAGVSENGVNLRLDVY----RTTGAFDANGKRIAPPEAQTD 234
DB 262 LTRLSGDLGGPQQYFAG-----SLGALDYSFDGTRHVGASYDAHGRIAPESQGD 315
QY 235 KQDSKSLSVNTVNDWQDKQINLALTHYNDKQDTPDYADPDYGNRLAVLFGKPSLNAIK 294
DB 316 LFDNSNYNIGGKLGILGRVINGNORQLALSHYDARQDTPDYATD--PRVARLPPGSGVPAANAII 373
QY 295 GLSLSPQTKTSTFNINHHDDLMGNTNTNAYYRREKGRFYFPFVAPFSIAKALPILQS 354
DB 374 GLELDEQNRINTLANLEYNDILGSRISAQLYRYDYTRTFPDA----- 420
QY 355 MNLPSATLDAYTKAPOARAYGVQLQSKSKEAELVGRVPLNKP 395
DB 421 -----RAVSTRGNVDQIQWQNSEVFGSRILTTRTP 449

RESULT 2
Q9HVC0
ID Q9HVC0 PRELIMINARY; PRT; 742 AA.
AC Q9HVC0;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Probable TonB-dependent receptor.
GN PA4675.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Golltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saler M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AB004881; AAG08062.1; -.
DR InterPro; IPR000847; HTH_LyER.
DR InterPro; IPR002011; RTKinaseII.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS00044; HTH_LYXR_FAMILY; UNKNOWN_1.
DR PROSITE; PS00239; RECEPTOR_TYR_KIN_II; UNKNOWN_1.
KW Receptor; Complete proteome.
SQ SEQUENCE 742 AA; 80952 MW; 222AA0A57E2F250A CRC64;

```

Query Match 24.6%; Score 503.5; DB 16; Length 742;  
 Best Local Similarity 30.8%; Pred. No. 7.9e-25;  
 Matches 132; Conservative 80; Mismatches 171; Indels 45; Gaps 10;

```

QY 11 SLPLLSVATVQQLYQAPNSLPTVELEPVVITIDKSGMALANRITQMPHTTKVYEEQIQ 70
DB 4 SIPLPAPLALSLSLPASFAALAADPFYQQMVVIGSRAPTSISLPGTVVWVIREQID 63
QY 71 EQATGSRQLADVAQQLPSLGVSSGTTSNFGQTMHGRQVQVFLNGVPLTGRSDISRLQNS 130
DB 64 QOTQAGVPLKEALQQLIPGLDITGSGQRTNNGQNLGRSVLWIDGVSINSSRGISRPQDS 123

```

```

QY 131 INPNOVARIEVLGATSIYSGATGGLINIVTKSDL-BEEQFETRIGVHGSKLSEGGIGY 189
DB 124 IDPFNIEIRIEVMSGSAVYGGATGGIINIVTKKGVGDTRFNTFELGARSGFQSHEDHL 183
QY 190 QVQSVAGVSENGVNL--ARLDVDYRTTGGAFDANGKRIAPPEAQTKQDSKSLSVNTNV 247
DB 184 RAAQSISG-----GNDLFNGRLAIAVQKNGAAYDGGDQVLTDTQTDLQYNRSDVLMGSL 239
QY 248 DWQLDDKQINLALTHYNDKQDTPDYADPDYGNRLAVLFGKPSLNAIK-GLSLSEQPKTTK 306
DB 240 GFTFANGHSLDLGLQYDSDGYDGRGLDLGNFDALRGAP--YSIKGVLDLDRPESEKR 297
QY 307 STFNINYYHDDLMGNTNTNAYYRREKGRFYFP-----VAPFSIAK----- 347
DB 298 HQFNATYHAPEVLGHDLVLAQYIRNEKMAFNFPPTIRYSNTGAINYGTSTYSASQDDTDY 357
QY 348 ---ALPILQSMNLPSATL-----DAYF-----KAPQARAYGVQLQSKSKEAELVGRVPLN 392
DB 358 YGMKLALVKTVERASLTGYVDLDRKFTSDOMLFNLPAAASGGLVASEQAK-LGRYPDI 416
QY 393 NKPKEALF 400
DB 417 DTDSRAFF 424

RESULT 3
Q8XSQ1
ID Q8XSQ1 PRELIMINARY; PRT; 812 AA.
AC Q8XSQ1;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Probable ferric siderophore receptor outer membrane signal peptide
DE protein.
GN RSP0416 OR RS00873.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMI1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choienne N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguiet P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RA "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646078; CAU17567.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
KW Plasmid; Complete proteome.
SQ SEQUENCE 812 AA; 87434 MW; DB03190282004BF3 CRC64;

```

Query Match 24.3%; Score 497; DB 16; Length 812;  
 Best Local Similarity 35.8%; Pred. No. 2.4e-24;  
 Matches 114; Conservative 65; Mismatches 123; Indels 16; Gaps 5;

```

QY 29 ESLPTVELEPVVITIDKSGMAL---ANRITQMPHTTKVYEEQIQEQATGSRQLADVAQ 85
DB 112 EAPAAVELEAVTWSDTRSDLRPRTRVQTQISH-----EELDQAGSGTLATALSK 164
QY 86 LIPSLGVSSGTTSNFGQTMHGRQVQVFLNGVPLTGRSDISRLQNSINPNOVARIEVLGSA 145
DB 165 VVPGMADSSHTITDYGQTLRGKGTLLVDGMPLNTRNDRSARNLASLDPNRIEVLGRS 224
QY 146 TSIYSGATGGLINIVTKSDLEEFQFETRIGVHG--SKLSSEGIGYQVQSVAGVSENGN 203

```

```

Db 225 SALKYSGATGAGVISTTRPAGGPRRAETVTSMTSPUSHGCAAGLGGAAQOYVAG--SKGP 282
Qy 204 VLARLDVDRRTTGAFADPANGKRIAPAPACTDKODSKLSNTNVMDQLDKONINLALH 263
Db 283 VYEFVFNQGHVGGSDADAGHRIAPPSOGDLFDSNITYSGKLGKRIKIPNQVYVSASH 342
Qy 264 YNDKQDTDYAPDYGNRLAVLFGEKPSLNAIKGLSLSEOPRTKSTFNINHHDDLWGNTI 323
Db 343 YDAKQDTDYASD--PEVALPLPAGSAVARAINGLQADQNRIRKTLNLQYENMDVFGSQV 400
Qy 324 NNNAYYRREKGRFPYFVA 341
Db 401 GAQFYRYDRFSRPAPDA 418

RESULT 4
ID 0934J5 PRELIMINARY; PRT; 725 AA.
AC 0934J5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Ferric aerobactin receptor (ferric aerobactin receptor precursor
  Iuta).
GN IUTA.
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxId=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MP1;
RA Funahashi T., Yamamoto S.;
RT "Cloning and characterization of the iuta gene in Vibrio
  parahaemolyticus."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB050700; BAB47148.1; -
DR EMBL; AB066099; BAB83802.1; -
DR InterPro; IPR000531; TONB_boxC.
DR Pfam; PF00593; TONB_boxC; 1.
KW Receptor.
SQ SEQUENCE 725 AA; 79329 MW; 57BD1CE14CDA874B CRC64;

Query Match
Best Local Similarity 33.2%; Score 493; DB 2; Length 725;
Matches 136; Conservative 68; Mismatches 156; Indels 50; Gaps 10;

Qy 15 LSAVATQQLYAQ--PRESLPTVELBEPVITIDSGMALNRITQMPHTTKVIEEOIOEO 72
Db 24 LLAABQQAOLASQAADQLVYV-----GSVMPKSIDIDIGTWVFGQSEIAQ 71
Qy 73 ATGSRDLADVMAQLISLGVSGTTSNFGQTMHGRQVQLANGVPLTGRDLSROLNSTIN 132
Db 72 YRAGKTLGILSATIPSLDVGTRGTYGQNLGRAMLWIDGVSLQSSRPISROLDAID 131
Qy 133 PNQVATIEVLSGATISYGSGATGLINIVTKSLDEE--OPEIRIGVHGSLSSEGIGYQV 191
Db 132 PPRIRIEIVLSGATISYGAGATGVNITIKKAYSDELAESEFVGTSGRNSSDDPDYKV 191
Qy 192 GGSVAVSENGVNLARLDVDRRTTGAFADPANGKRIAPAPACTDKODSKLSVNTNVMDQL 251
Db 192 AGQVAG--GNDIVKARGSVYSETQAFDGNQDVIYIPDISQSLQYNSTLIDVWGSALQI 249
Qy 252 DDKONINLALTHYNDKQDTDYAPDYGNRLAVLFGEKPSLNAIKGLSLSEOPRTKSTFNTI 311
Db 250 SDASKINLVAQYVDSQDSFY-----GLYIVNSKFDVVRKGFSDRHEGTERVLLSA 301

```

```

Qy 312 NYHDDLWNGTINTINAYYRERKGRFPYFVAPPS-----IAKALPILQSNLPSAT--LD 363
Db 302 NYAHNDNVFQHQILGSLSTRSEDDQTFTPYQSSSQGETVEFARLALASWGFSAVYGD 361
Qy 364 AY-----TKAPQARAYGLOSEKAEVYLGVRPNINKPRALF 400
Db 362 AYLDKDFDSNQALFDXTIADNS---GNLNRITYAEV--GRYGVDTVYSYAMF 407

RESULT 5
ID 08ZHB0 PRELIMINARY; PRT; 726 AA.
AC 08ZHB0;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative ferric siderophore receptor.
GN IUTA OR YP00994.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
  Yersinia.
OX NCBI_TaxId=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
RX MEDLINE=21470413; PubMed=11586360;
RA Parhill J., Wren B.W., Thomson N.R., Tilball R.W., Holden M.T.G.,
  Prentice M.B., Sedahlia M., James K.D., Churcher C., Mungall K.L.,
  Baker S., Basham D., Bentley S.D., Brooks K., Gerdon-Tarraga A.M.,
  Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
  Fellwell T., Hamlin N., Holroyd S., Jagels K., Katlyshev A.V.,
  Leather S., Mouton S., Oyston P.C.F., Quail M., Rutherford K.,
  Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527(2001).
DR EMBL; AJ414145; CAC89837.1; -
DR InterPro; IPR000531; TONB_boxC.
DR Pfam; PF00593; TONB_boxC; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.
KW Receptor; Hypothetical protein; Complete proteome.
SQ SEQUENCE 726 AA; 80375 MW; 4EF72CAB010EF94C CRC64;

Query Match
Best Local Similarity 30.7%; Score 491; DB 16; Length 726;
Matches 123; Conservative 69; Mismatches 155; Indels 54; Gaps 9;

Qy 1 MESHYPQWLSPLLSAVATQQLYAQPNESLPTVELBEPVITIDSGMALNRITQMPHT 60
Db 1 MKKHKL--WVNLPCLLVMLTPAAMAR-----DQLVGSANSHRSVA---EMAQT 44
Qy 61 TKYIYBEQIQEQAQTSRQADVMAQLIPSLGVSGTTSNFGQTMHGRQVQLANGVPLTG 120
Db 45 TWVIEQBLEQVOVGLEKIDILQILPGIDVSSQRTYVGNMGRKRSIMWIMIDGVRILNS 104
Qy 121 SRDISROLNSINPNQVATIEVLSGATISYGSGATGLINIVTKSLDEEBOFPRIGVHS 180
Db 105 SRDSRQLOSDIPFNIAHLEIVSGATSLVGGSTGLINIVTKKQEGQVEQLQIGKRG 164
Qy 181 KLSSEGIQVQSVAGVSENGVNLARLDVDRRTTGAFADPANGKRIAPAPACTDKODSKS 240
Db 165 FNSHNDHENISAMSGGTE--RAGRFVSVYQRGGMWDGKNEVLLINTQGLQYSNR 222
Qy 241 LSVNTNVMDQLDKONINLALTHYNDKQDTDYAPDYGNRLAVLFGEKPSLNAIKGLSLSE 300
Db 223 LDVMGTGLINIDENQQLQTLTQYFNSESQKGLYIGQNFSAVTGQASNS--AALNSDR 281
Qy 301 OPKTTSTFNINHHDDLWNGTINTINAYYRERKGRFPYFVAPPSIAKALPILQ----- 353
Db 282 IPTEBHLINLQYSNTDFGQDVAQVYTRDESILTFYP-----PTLKDGQVSTI 331
Qy 354 -----SMNLPSATLDAYTKAPQARAYGV--LQSES 381

```

Db 332 GASQKTDYFGSKLTNSPIDSULT-----LTYGIDLEHES 367

RESULT 6

Q9XCHO PRELIMINARY; PRT; 732 AA.  
 AC Q9XCHO;  
 DT 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE Iuta.  
 GN IUTA.  
 OS Shigella flexneri.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Shigella.  
 OC NCBI\_TaxID=623;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=M90T;  
 RX MEDLINE=99340540; PubMed=10411725;  
 RA Moss J.E., Cardozo T.J., Zychlinsky A., Groisman E.A.;  
 RT "The selC-associated SHI-2 pathogenicity island of Shigella  
 flexneri."  
 RL Mol. Microbiol. 33:74-83 (1999).  
 DR EMBL; AF141325; AAD4750.1; -.  
 DR InterPro; IPR000531; TonB\_boxC.  
 DR Pfam; PF00593; TonB\_boxC; 1.  
 DR PROSITE; PS01156; TONB DEPENDENT REC 2; UNKNOWN 1.  
 SQ SEQUENCE 732 AA; 80984 MW; F3A1E0A1B7473210 CRC64;

Query Match 23.1%; Score 472; DB 2; Length 732;  
 Best Local Similarity 30.9%; Pred. No. 9.1e-23;  
 Matches 130; Conservative 65; Mismatches 176; Indels 50; Gaps 9;

QY 10 LSLPLLSVAVTQQLYAQPNESLPTVELEPVITDKSGMALANRITOMPHTTKVIYEEQI 69  
 Db 14 LLLVLSPAVAQO--NDDNE-----IIVCASRSNRITVA-----EMAQTWVIENAE 58  
 QY 70 QEQATGSRQLADVMAQLIPSLGVSSGTTNFGQTMHGRQVQLNGVPLTGSRDISRQLN 129  
 Db 59 EQQIGGKELKDALAQLIPGLDVSSQSRNTNYGMNMRGPLVLVLDGVRLNSRSRQ 118  
 QY 130 SINPNQVARIIEVLGATSIYGSATGGLINIVTKSDLEEQEPETRIQVHSGKLSSE 185  
 Db 119 SVDPNIDHIEVISGATLYGGSTGGGLINIVTKGQPTMMEFEAGTKSGFNSSKHDE 178  
 QY 190 QVGSQVAGVSENGNVLARDVDYRTGGAFDANGKRIAPEPAQTDKQDSKLSVNTVDW 249  
 Db 179 RIAGAVSG--GNDHISGRLSVAYQKFGWFDGNGDATLIDNTQTGLQHSNRDL 236  
 QY 250 QLDKQKINLALTHYNDKQDTPADYDGNRLAVLFGK-PSLNAIKGLSLSEQPKTKST 308  
 Db 237 NIDESRQLQITQYKSGDDNYGLNGKFSASISGSTPYVS--KGLNSDRIPGTERHL 294  
 QY 309 FNIYHHDLWGNTINTNAYRREKGRFPFVAPFSAIAKALPILOS-----MNL 358  
 Db 295 ISLQYSDSDFLGQELVGQVYRDESLRYPPFTVNANKQATAFSSSQDQDQYGMKLT 354  
 QY 359 SATLDA-----YTKAPQARAYGVLOSKEAEVLGRVPLNPKRAL 399  
 Db 355 SQLMDCGQITWGLDAEHERFTSNQMFDDLAQASASGGLNNH-KIYTTGRYP 413  
 QY 400 F 400  
 Db 414 F 414

RESULT 7

Q93PE2 PRELIMINARY; PRT; 732 AA.  
 AC Q93PE2;  
 DT 01-DEC-2001 (Tremblrel. 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)

DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
 DE Iuta.  
 GN IUTA.  
 OS Shigella boydii.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Shigella.  
 OC NCBI\_TaxID=621;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=0-1392;  
 RX MEDLINE=21311742; PubMed=11418557;  
 RA Purdy G.E., Payne S.M.;  
 RT "The SHI-3 Iron Transport Island of Shigella boydii 0-1392 Carries the  
 Genes for Aerobactin Synthesis and Transport."  
 RL J. Bacteriol. 183:4176-4182 (2001).  
 DR EMBL; AF335540; AAK71635.1; -.  
 DR InterPro; IPR000531; TonB\_boxC.  
 DR Pfam; PF00593; TonB\_boxC; 1.  
 DR PROSITE; PS01156; TONB DEPENDENT REC 2; UNKNOWN 1.  
 SQ SEQUENCE 732 AA; 81013 MW; 57A3D10AEF70AE31\_CRC64;

Query Match 23.1%; Score 472; DB 2; Length 732;  
 Best Local Similarity 31.1%; Pred. No. 9.1e-23;  
 Matches 133; Conservative 62; Mismatches 170; Indels 62; Gaps 11;

QY 10 LSLPLLSVAVTQQLYAQPNESLPTVELEPVITDKSGMALANRITOMPHTTKVIYEEQI 69  
 Db 14 LLLVLSPAVAQO--NDDNE-----IIVCASRSNRITVA-----EMAQTWVIENAE 58  
 QY 70 QEQATGSRQLADVMAQLIPSLGVSSGTTNFGQTMHGRQVQLNGVPLTGSRDISRQLN 129  
 Db 59 EQQIGGKELKDALAQLIPGLDVSSQSRNTNYGMNMRGPLVLVLDGVRLNSRSRQ 118  
 QY 130 SINPNQVARIIEVLGATSIYGSATGGLINIVTKSDLEEQEPETRIQVHSGKLSSE 185  
 Db 119 SVDPNIDHIEVISGATLYGGSTGGGLINIVTKGQPTMMEFEAGTKSGFNSSKHDE 178  
 QY 186 GIGYQVGSQVAGVSENGN--VLARDVDYRTGGAFDANGKRIAPEPAQTDKQDSKLSV 243  
 Db 179 -----RIAGAVSGGNDHISGRLSVAYQKFGWFDGNGDATLIDNTQTGLQHSNR 230  
 QY 244 NTNVDWQLDKQKINLALTHYNDKQDTPADYDGNRLAVLFGK-PSLNAIKGLSLSEQ 302  
 Db 231 MGTGLNIDESRQLQITQYKSGDDNYGLNGKFSASISGSTPYVS--KGLNSDRIP 288  
 QY 303 KTKTKTFNINHHDLWGNTINTNAYRREKGRFPFVAPFSAIAKALPILOS----- 354  
 Db 289 GTERHLISLQYSDSDFLGQELVGQVYRDESLRYPPFTVNANKQATAFSSSQDQDQY 348  
 QY 355 --MNLPSATLDA-----YTKAPQARAYGVLOSKEAEVLGRVPLN 393  
 Db 349 MKLTNSQLMDCGQITWGLDAEHERFTSNQMFDDLAQASASGGLNNH-KIYTTGRYP 407  
 QY 394 KPKRALF 400  
 Db 408 ITXLA 414

RESULT 8

P72594 PRELIMINARY; PRT; 851 AA.  
 AC P72594;  
 DT 01-FEB-1997 (Tremblrel. 02, Created)  
 DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)  
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
 DE Ferric aerobactin receptor.  
 GN IUTA OR SL1206.  
 OS Synechocystis sp. (strain PCC 6803).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.  
 OC NCBI\_TaxID=1148;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97061201; PubMed=8905231;

RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,  
 RA Miyajima N., Hiroseawa M., Sugita M., Sasamoto S., Kimura T.,  
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naro K., Okumura S.,  
 RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,  
 RA Tabata S.,  
 RT "Sequence analysis of the genome of the unicellular cyanobacterium  
 RT *Synechocystis* sp. strain PCC6803. II. Sequence determination of the  
 RT entire genome and assignment of potential protein-coding regions.";  
 RL DNA Res. 3:109-136(1996).  
 DR EMBL; D90899; BAA16594.1;  
 DR InterPro; IPR001202; WW\_Rsp5\_WMP.  
 DR PROSITE; PS01159; WW\_DOMAIN\_1;  
 KW Complete proteome.  
 SQ SEQUENCE 851 AA; 91657 MW; F5AC182FC149ACD CRC64;

Query Match 22.2%; Score 453.5; DB 16; Length 851;  
 Best Local Similarity 33.0%; Pred. No. 1.9e-21;  
 Matches 113; Conservative 67; Mismatches 143; Indels 19; Gaps 10;

QY 22 QLYAQPNESLPTELE-----PVVITIDKSGMALNRITOMPHTTKYIEBOIQEATGS 76  
 DB 153 QULIRGVAGTTEIESEMLQIVSATRTREBIAN-----IPRSVYIEBAIEQOTQY 208  
 QY 77 ROLADYMAQILPSLGVSSGTTNFGQTMGRQYQFLINGVPLTGSDDISQINSINPNOY 136  
 DB 209 SSIALDILGOLVPLGAPSTGSASQFGALRGRLVLDGVQFTNRNAPRDLQTIAPSAI 268  
 QY 137 ARLEVLSGATSIYSGATGGLINIVYKSLDEEOP--ETRIGHGSKLS--SEGIGYQVGS 194  
 DB 269 ERLEVIQGPALYIGDAGTGVINIIITRGAPPEPLASTRAIINTDPSVNSLGRWVEQY 328  
 QY 195 VAGVSENGVNLARLDVYRTTGAFFDANGKRIAPEP--AQTDKDSKLSLVNTVMDQLD 253  
 DB 329 VGGTLDYADY--AFYASVEYVGCFDNLGRIRIPDPGQGVSDTDSFVNLGKLGIMTD 386  
 QY 254 KQINIALTHYNDKQDTPADYDGNRLAVLFGEKPSINAIKGLSLSEQPTTKST--FNIN 312  
 DB 387 EQLQITINHFQATQNTDFTVD--PSITAIAGHRS--QALIDGLD--DTPQTSNNVTYVSLD 442  
 QY 313 YHHDDLGMGNTINTNAYYRREKGRFPYVAPFSAKALPIQS 354  
 DB 443 YSHSNVANGNLKGQIYRDYLTFRFPDGRASVLSGNSIFQS 484

## RESULT 9

Q9YZR0 PRELIMINARY; PRT; 867 AA.  
 AC Q9YZR0;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical protein A110397.  
 GN ALK0397.  
 OS *Anabaena* sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OX NCBI\_TaxID=103690;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21595285; PubMed=11759840; Kuritz T., Sasamoto S.,  
 RA Kaneko T., Nakamura Y., Wolk C.P., Kawashima K., Kimura T.,  
 RA Watanabe A., Iriuch M., Ishikawa A., Kawanishi K., Muraki A.,  
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Yamada M.,  
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yasuda M.,  
 RA Yasuda M., Tabata S.,  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 RT cyanobacterium *Anabaena* sp. strain PCC 7120.";  
 RL DNA Res. 8:205-213(2001).  
 DR EMBL; AEO03582; BAB72355.1;  
 DR InterPro; IPR002106; AATRNA\_LigaseII.  
 DR InterPro; IPR000531; TonB\_boxC.  
 DR InterPro; IPR001202; WW\_Rsp5\_WMP.  
 DR Pfam; PF00593; TonB\_boxC.1.  
 DR PROSITE; PS00339; AA\_TRNA\_Ligase\_II\_2; UNKNOWN\_1.

DR PROSITE; PS01159; WW\_DOMAIN\_1;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 867 AA; 94332 MW; CE812BD0E78B537F CRC64;

Query Match 22.1%; Score 453; DB 16; Length 867;  
 Best Local Similarity 33.5%; Pred. No. 2.1e-21;  
 Matches 110; Conservative 63; Mismatches 143; Indels 12; Gaps 6;

QY 14 LLSVANTQQLVAPNESLPTELEPVVITIDKSGMALNRITOMPHTTKYIEBOIQEATGS 73  
 DB 165 ILAFITRBAQSPAEVDEPIEL--VTRATREPIQN-----VPSSTIVDEBOIAQA 217  
 QY 74 TGSROLADYMAQILPSLGVSSGTTNFGQTMGRQYQFLINGVPLTGSDDISQINSINP 133  
 DB 218 STSRNLIETLKGTPGLAPAGASFGTLGRNROVLIDGVPOSTTNASRLATIDA 277  
 QY 134 NOYARLEVLSGATSIYSGATGGLINIVYKSLDEE--QETRIIGHGSKLSSEGIGYQV 192  
 DB 278 ALIERIEVRGSPALYIGDAGTGVINIIITRRPTEKLTSTRTEGVSAALGNLEGDSFSTN 337  
 QY 193 QSVAGVSENGVNLARLDVYRTTGAFFDANGKRIAPEP--AQTDKDSKLSLVNTVMDQL 251  
 DB 338 LGHFTSAKQGNDFTFNPAVAKNGFFDAGDRIPSDPAQGGAPDASSINLFGKFGIDI 397  
 QY 252 DQKINIALTHYNDKQDTPADYDGNRLAVLFGEKPSINAIKGLSLSEQPTTKSTFNI 311  
 DB 398 DANQRLQITNHFQATQNTDFTVD--PRVNTIPGRKA--RALEGSLDPRPGNENTFINL 454  
 QY 312 NYHDDLGMGNTINTNAYYRREKGRFPY 339  
 DB 455 QYTHDDLFNASKLQALYRDYLTFRFPF 482

## RESULT 10

Q9ZNM3 PRELIMINARY; PRT; 705 AA.  
 AC Q9ZNM3;  
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE IUTA.  
 GN IUTA.  
 OS *Vibrio orientalis*.  
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; *Vibrio*.  
 OX NCBI\_TaxID=28175;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SD004;  
 RA Murakami K., Fuse H., Takimura O., Inoue H., Yamaoka Y.,  
 RT "Cloning and characterization of iuta gene from *Vibrio* species.";  
 RL Submitted (Feb-1998) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AB010890; BAA74703.1;  
 DR InterPro; IPR000531; TonB\_boxC.  
 DR Pfam; PF00593; TonB\_boxC.1.  
 DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; UNKNOWN\_1.  
 SQ SEQUENCE 705 AA; 77907 MW; 4f0CB4C074E44212\_CRC64;

Query Match 21.5%; Score 441; DB 2; Length 705;  
 Best Local Similarity 33.7%; Pred. No. 9.4e-21;  
 Matches 115; Conservative 67; Mismatches 129; Indels 30; Gaps 8;

QY 11 SLPILSVAV-----TQQLVAPNESLPTELEPVVITIDKSGMALNRITOMPHTTKYI 65  
 DB 7 SFLPSTILAGVAATVPAQAQOEY--TSEERKVVVS--SHMPKA--ISDIPGTWYID 59  
 QY 66 BEQIQEATGSROLADYMAQILPSLGVSSGTTNFGQTMGRQYQFLINGVPLTGSDDIS 125  
 DB 60 SETIBOEVYGGSLDIILATIPSLDVSSGGRTHSQNLGRGSLMMIMIGVSLQSVRSIS 119  
 QY 126 RQINSINPNOYARLEVLSGATSIYSGATGGLINIVT--KSLDEEOPETRIIGHGSKLS 184  
 DB 120 RQDSIDIPNIDRIEVLVSATSIYGGASGAGVINITTAKASGEVEFVSQTSVSSS 179

```

QY 185 EGIGYQGVGVSENGNVLARLDVYRTGGAFDANGKRIAPPAQTDKQDSKLSVN 244
Db 180 EDPDYKLAQSIAGNE--KVQGRASVAYTKTQGYFDAGGEMVTPDITQGSTQFNETIDL 237
QY 245 TNDVMDLDDKQNLINLTHALYNDKQDTPADYDYNRLAVLFGKPS-----LNAIKGLSL 298
Db 238 GNLQINIAEGQQLNLAQYDSQDDTPYG-----LYFKDGGSGNFQFVDVREGYSA 288
QY 299 SEQPKTKSTFNINHHDLWGNNTINTNAYYRREKGRFPVF 339
Db 289 DREQGTERVMSLAAYNNFLSHLIAELSYRTEDHTFMPY 329

RESULT 11
ID Q8YT2 PRELIMINARY; PRT; 880 AA.
AC Q8YT2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Hypothetical protein Alr2581.
GN ALR2581.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003589; BAB74280.1; -;
DR InterPro; IPR001064; Crystallin.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 880 AA; 96442 MW; 790955942CE4BFF8 CRC64;

Query Match 19.6%; Score 400.5; DB 16; Length 880;
Best Local Similarity 30.3%; Pred. No. 68-18;
Matches 103; Conservative 58; Mismatches 134; Indels 45; Gaps 10;

QY 17 VAVTQQLYAPNESLPTVELEPVVITDKSG-----NALANR---ITOMPHTTKVIYEQ 68
Db 159 VVSTTQTPTQEQ--PTSETPDATTSESDSSIELVVTATRTREEDIQNVPRTVITREQ 216
QY 69 IQEQATGSRQLADVMAQLIPSLGVSSGTTSNFGQTMHGRQVQFLNGVPLTGS-RDISRQ 127
Db 217 IEDQARLSTNLADILAKTPVGFSGPSIRDTDTFCQNLGRNISVLIDGVFPQGNLQSFSAQ 276
QY 128 LNSINPNQVARIIVLGSATSIYSGATGLINIVTKSDLEER-QFTRIGVHGSKLSSP- 185
Db 277 LTTIDPSAIEVRVGNPAIYGGQATGGVNIITKRPSCQKLTSTNIGLDTSLTRSD 336
QY 186 GIGYQGVGVSENGNVLARLDVYRTGGAFDANGKRIAPPAQTDKQDSKLSVN 245
Db 337 SFGYNLSHQIAGT--EGKEDYTVGFLSVLTAGFYDAGRIANFAGD---DSTKINALA 391
QY 246 NVDWLDDKQNLINLTHYNDKQDTPADYDYNRLAVLFGKPSLNAIKGLSLSEOPK-- 303
Db 392 KIGVELSPQRLQFTFNHFNQONSDFISD-----PAVDDIPGIQKSRALKLP 439
QY 304 -----TTKSTFNINHHDLWGNNTINTNAYYR 330
Db 440 EGTTVIGADQGISITNTLLSLNNYNNNDNIFGSKLQQAQYR 479

RESULT 12
Q8YT3

```

```

ID Q8YT3 PRELIMINARY; PRT; 536 AA.
AC Q8YT3;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Ferric aerobactin receptor.
GN ALR2620.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003590; BAB74319.1; -;
KW Receptor; Complete proteome.
SQ SEQUENCE 536 AA; 58730 MW; 3CC3613C03F1B4E1 CRC64;

Query Match 16.3%; Score 333.5; DB 16; Length 536;
Best Local Similarity 23.8%; Pred. No. 7.3e-14;
Matches 93; Conservative 79; Mismatches 148; Indels 71; Gaps 10;

QY 11 SLPLLSVAVTQQLYAPNESLPTVELE---PVVITDKSGNALANRITOMPHTTKVIYEE 67
Db 152 SVFPQPPQPTQEQPTQEQTPQPSAEVDDAIELIVTATTEA---ITNVPRTVITRE 207
QY 68 IQEQATGSRQLADVMAQLIPSLG-VSSGTTSNFGQTMHGRQVQFLNGVPLTGSRDISR 126
Db 208 EIEKQSTITNDLGDILGRTVFGLPNSLNAGNATURGPVSLIDGVFQGGNSFVNT 267
QY 127 QNLSINPNQVARIIVLGSATSIYSGATGLINIVTKSDLEEEQFETRIGVHGSKLSE 186
Db 268 QLEYISPAIERVEVVRGPTAVFGQSGGVNIITKRPAGFTSTAQVGISAA----- 321
QY 187 TGYQGVGVSENGNVLAR-----LDVYRTTGGAFDANGKRIAPPAQTD 234
Db 322 ---AAGDAFLGNSFGNYLQYFGSKDGIIFYVFSLSNSVGGFFDAGSRIPSNATSD 378
QY 235 KQDSKLSLVNTNVDLDDKQNLINLTHYNDKQDTPADYDYNRLAVLFGKPSLNAIK 294
Db 379 --DTVSTNIGKIGIDVGEQQLQFTVNHGNSRKIKFIAD-----PITRTIP 424
QY 295 GLSL-----SEQPKTKSTFNINHHDL-----WGNNTINTN 326
Db 425 GLOTTRALRQNVVEGTDAPIRTSTSVNLVTTIPSPILEGELMYKLPFLRSWAKTISP 484
QY 327 AYRREKGRFPVFVAPPSIAKALPILQSNVL 357
Db 485 NSFQYRKAFFGLVLTQYQGFQSLTVLVSSRL 515

RESULT 13
Q8YT2
ID Q8YT2 PRELIMINARY; PRT; 882 AA.
AC Q8YT2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Ferric aerobactin receptor.
GN ALR2209.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;

```

RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,  
 RA Watanabe A., Iritaguchi M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shimizu S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.,  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 RT cyanobacterium Anabaena sp. strain PCC 7120."  
 RL DNA Ref. 8:205-213(2001).  
 DR EMBL; AP003588; BAB73908.1; -  
 DR InterPro; IPR002106; AARNA\_1lgaseII.  
 DR InterPro; IPR000531; TonB\_boxC.  
 DR Pfam; PF00593; TonB\_boxC; 1.  
 DR PROSITE; PS00339; AA\_TRNA\_LIGASE\_II\_2; UNKNOWN\_1.  
 KW Receptor; Complete proteome.  
 SQ SEQUENCE 882 AA; 96731 MW; 4CB94080440CFD9C CRC64;

Query Match 16.0%; Score 326.5; DB 16; Length 882;  
 Best Local Similarity 27.8%; Pred. No. 4,4e-13;  
 Matches 91; Conservative 60; Mismatches 145; Indels 31; Gaps 11;

QY 27 PNESLPYVLEEVVITIDSG---MALNR---ITOMPHTKVYEQIOEQATGSRQ 76  
 DB 170 PPTPTSTBEOPTABDDPIELVYATREBDIQNVPRSVYITRQEQOTTVRD 229  
 QY 79 IADVAQLPSLGVSSTGTSNFGQTHGROVQFLNGVPLTGRDIS---ROLNSINPQ 135  
 DB 230 LTIILGNVTVGLGASAESQSFQTLRGRPLILVDGVPISSINDIDTSVNLRRIDVGA 289  
 QY 136 VARIYLSATISYSGATGLINITY-KSDLEBEQFETRIYHG-SKLSSEIGYQVQ 193  
 DB 290 IRIEIVRRPSPAVYDGAAGVINIITRRPDQVVSNAEIGISVGNKSSFGFVAVY 349  
 QY 194 SVAGSENGVTLARLDVDRITGAFDANGKRIAPPAQTDKO-DEKSLSVNTNVDMOLD 252  
 DB 350 GISGQGGVDPIASFPRD--SGTPTDABGDRI--PLGDMEANASINVLGKGFQUG 404  
 QY 253 DKONINLALTHYNDKQDTPADYDGNRLAVLPGEKPSLNAIKGLSU-----SEQPKTK 306  
 DB 405 SSGRLQITANVYDQDSNV--DYD---LTVGQIRGIGIKARALDQVGFNSTNPFNG 458  
 QY 307 STEFNINHHDDLGMGTINTNATYRREX 333  
 DB 459 TVIQLDYTHNNTLSQLQAQAYYRQTK 485

## RESULT 14

Q44003 PRELIMINARY; PRT; 840 AA.  
 ID Q44003;  
 AC Q44003;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 GN Ferric alkaligin B.  
 OS ALB.  
 OS Ralstonia metallidurans.  
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
 OC Ralstonia.  
 NC NCB1\_TaxID=119219;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CH34;  
 RX MEDLINE=96404801; PubMed=8808942;  
 RA Gills A., Khan M.A., Cornelis P., Meyer J.M., Mergaey M.,  
 RA van der Leije D.,  
 RT "Strophote mediated iron uptake in Alcaligenes eutrophus CH34 and  
 RT identification of aeb encoding the ferric alkaligin B receptor."  
 RL J. Bacteriol. 178:5499-5507(1996).  
 DR EMBL; X97499; CAA66129.1; -  
 DR InterPro; IPR000531; TonB\_boxC.  
 DR Pfam; PF00593; TonB\_boxC; 1.  
 SQ SEQUENCE 840 AA; 91733 MW; 72A1F4E5257430ED CRC64;

Query Match 15.9%; Score 324.5; DB 2; Length 840;

Best Local Similarity 26.4%; Pred. No. 5.5e-13;  
 Matches 102; Conservative 68; Mismatches 132; Indels 85; Gaps 11;

QY 45 KSGMALNRIOTMPHT-----TKVIYEQIOEQATGSRQIADVAQOLPSLGVSSTTS 98  
 DB 146 RAGPGGSHRVVIDIERSGPPDAIDVLETELELKTGSDSLATVSKVIFGMADSHVT 205  
 QY 99 NFGQTHGROVQFLNGVPLTGRDISROLNSINNOVARIYLSGATSYSGATGGLI 158  
 DB 206 DFGQLRGRGMLVLLDGIPLMTNRNSANLNIDPALVERVFLNGSSAIVGCGPTGGIV 265  
 QY 159 NIVYTSDLSEBEQFETRIYHG-SKLSSEIGYQVQGVAVYSENGVNLAR-LVDYRTT 215  
 DB 266 SITTPAGCEPADTVMSTPTLSLRGRLECRPAALFGQGRARLCVRRAPAYQ- 324  
 QY 216 GGAFDANGKRIAPPAQTDKQDSKLSVNTNVDMOLDKONINLALTHYNDKQDTPAD 275  
 DB 325 --SYDAKGHALRRSQARA-----TCSTRT-----TTTSAASSAFAPD 360  
 QY 276 YGNRLAVLPGEKPS-----LNAIKGLSLSEQPKTKSTFNINHH 315  
 DB 361 QRIQASSAILPSTRAGRRIRAWPNCRRACPRRALDGLADQNLRIITLVLETEH 420  
 QY 316 DDLGNTINTNATYRREKGRFPVAPPSIAKALPILQSMNLPATLDATYAPQAPAYG 375  
 DB 421 RDLTSSQLSAGFYNNYYSRFPF-----DARAVV-----RCGN 455  
 QY 376 VLOSSEKAEVIG-----RVPNLNPKR 397  
 DB 456 VDAQMDSEVFSGSLTVKTP-LGRDKR 481

## RESULT 15

P95494 PRELIMINARY; PRT; 646 AA.  
 ID P95494;  
 AC P95494;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE TonB-dependent receptor Hmnr.  
 GN Hmnr.  
 OS Porphyromonas gingivalis (Bacteroides gingivalis).  
 OC Bacteria; CFB group; Bacteroidetes; Bacteroidales; Porphyromonadaceae;  
 OC Porphyromonas.  
 NC NCB1\_TaxID=837;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A7436;  
 RX MEDLINE=99426790; PubMed=10496872;  
 RA Simpson W., Wang C.Y., Mikolajczyk-Pawlinska J., Potempa J.,  
 RA Travis J., Bond V.C., Genco C.T.,  
 RT "Transposition of the Endogenous Insertion Sequence Element IS1126  
 RT Modulates Gingipain Expression in Porphyromonas gingivalis."  
 RL Infect. Immun. 67:5012-5020(1999).  
 DR EMBL; U87395; AAB47566.2; -  
 KW Receptor.  
 SQ SEQUENCE 646 AA; 73157 MW; A57320CC565DBB63 CRC64;

Query Match 10.0%; Score 205; DB 2; Length 646;  
 Best Local Similarity 25.5%; Pred. No. 2.7e-05;  
 Matches 82; Conservative 50; Mismatches 105; Indels 84; Gaps 15;

QY 25 AQPNSLPT--VELEPVYTIIDKSGMALNRIOTOMPHTKVYEQIOEQATGSRQIADV 82  
 DB 29 AQPPTIVSGNALDIDVT-----GSRTARLKDVPVPTKVKADQIKAIAPSS--FIDV 82  
 QY 83 MQLPLPSLGVSSTGTSNFGQTHGROVQ-----FLNGVPLTGRDISROLNS 130  
 DB 83 LGYIIPGIEF-----THGSRDQNLNAGPDESSITFLVDEBLISTGSTSGIDPER 132  
 QY 131 INPNOVARIYLSGATSYSGATGLINITYKSDLEBEQFETRIYHGSKLSSEIGY 189  
 DB 133 INPDIERIEVLRGASALYGSNAIGVINIITRT--AKDPF--RVSASARYDSRDGKY 188







```

Best Local Similarity 23.6%; Pred. No. 7.3e-05;
Matches 83; Conservative 64; Mismatches 126; Indels 79; Gaps 18;

QY 7 FOWLSPLLSAVTOOLYA--QPNESTLPTVELEPVVITIDKSGMALANRITQMPHTTKVI 64
Db 8 FIWL-ITVSVTGSINPLSAESTDDNGETIVVE-----STAQVLRQPGVSVI 55
QY 65 YEEQEQEATGSRQADVAQAQLPSLGVSGTTSNFGQTHGROVQF-----LNL 114
Db 56 TRDDIOKNPP-VNDLADIIRKM-PGVNLTSSNS--GTRGNRRQIDIRGMGPENTLVLID 111
QY 115 GVPLT-----GSRDISRQLNSINPQVARIIVLSG-ATSIYSGATGGLINIVTK 163
Db 112 GVPVTSRNVSVRWGERDTRGDTNWVPPVEMVERIEIRMGPAARYGSGAAGGVNIIITK 171
QY 164 SDLEBEQFETRIGVHSGKLSSEG-----IGYQVGQSVAGVSENGVNLARLDVYRT 214
Db 172 RPTNDWHGSLSYTNYPESKSGDTRGRNGLSGPLAGDTLSTMRLYGN-LNRTDAD--- 227
QY 215 TGGAFDANG-----KRIAPEPAOTDKODSKLSVNTVNDWOLDKQKQINLALTHYNDKQPT 270
Db 228 ---SMDINSAGTKNAAGREGVTNK-----DINSVFSWKMTPOQILDPEAGY--SRQGN 276
QY 271 DYAPDYGRLAVLFGCKPSLNAIKGLSLSEQPKTKSTFNINY--HHDDLWG 320
Db 277 IYAGDTQN-----SNSNAV-TKSLAQSGRETNRLRYQNLGTHNGIWG 318

RESULT 21
QBXWSO PRELIMINARY; PRT; 621 AA.
ID QBXWSO;
AC QBXWSO;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE Probable outer membrane receptor protein.
GN RSC2400 OR RS02718.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=GM11000;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choiane N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646070; CAD16107.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Complete proteome.
SQ SEQUENCE 621 AA; 67468 MW; C715AA06C901BC4E CRC64;

Query Match 9.6%; Score 196.5; DB 16; Length 621;
Best Local Similarity 23.5%; Pred. No. 9.2e-05;
Matches 83; Conservative 55; Mismatches 136; Indels 79; Gaps 13;

QY 31 LPTVELEPVVITIDKSGMALANRITQMPHTTKVIIEEQEQEATGSRQADVAQAQL-IPS 89
Db 27 MPVGEINPTVVTASRSEQKAD---ALPHTTVISRADIERSQAPDAVSLRREAGIEIAQ 83
QY 90 LGVSSGTTSNFGQTHGROVQFLNGVPLTGRDTSRQLNSINPQVARIIVLSG-ATSI 148
Db 84 AGPGGAASLFRGAGSNQTLILIDGVRSVSGTGTGTQTEQLMAQDIDHIEIVRGVNSAL 143
QY 149 YGSGATGGLINIVTKSDLEEQF---ETRIGVHSGKLSSEGIGYQVGQSVAGVSENGVNL 205

```

```

Db 144 YCSDAIGGVQIFTRNGRHGAPLANABEYGARNTKRAQAGI-----SGOLD 190
QY 206 ARLDQVY-----RTTGGAFDANGKRIAP-----EPAQTDKQDSKLSVNTVNDWQ--- 250
Db 191 ARGDTSFALSALKTTG--FSSINPKQAPNPNPNPNAYANKSVSAQLQHRFSSDQAGL 248
QY 251 -----LDDKQIN--LALTHYNDKQDTPADYDYGRLAVLFGCKPSL 290
Db 249 TWFQWTSTVSYDNFAGPTDDNESHQVRAMSAVYDGLT---PWKTRLTLSQDGDKNL 305
QY 291 NAIKGLSLSEQPKTKS-----TFNINVHHDDLGMGTINTWAY 328
Db 306 NPTNGQVQBPGRFTRNRQASQWQNDWAFDPQMLKVGEHLN---QTIDTDAY 355

RESULT 22
Q93K73 PRELIMINARY; PRT; 725 AA.
ID Q93K73;
AC Q93K73;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Iron protein.
GN IRON.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=536;
RX MEDLINE=21295072; PubMed=11401961;
RA Dobrindt U., Blum-Oehler G., Hartesch T., Gottschalk G., Ron E.Z.,
RA Funfstuck R., Hacker J.;
RT "S-Fimbria-Encoding Determinant sfa(I) Is Located on Pathogenicity
RT Island III(536) of Uropathogenic Escherichia coli Strain 536.";
RL Infect. Immun. 69:4248-4256(2001).
DR EMBL; X16664; CAC43424.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.
SQ SEQUENCE 725 AA; 79349 MW; 585832AFA5BFF45 CRC64;

Query Match 9.6%; Score 196.5; DB 2; Length 725;
Best Local Similarity 24.3%; Pred. No. 0.00012;
Matches 83; Conservative 59; Mismatches 136; Indels 63; Gaps 16;

QY 9 WLSPLLSVAVTQQL-----YAQPNESLPTVELEPVVITIDKSGMALANRITQMPHTTKVI 64
Db 8 W-SLTVLLVGLNSQVSAKYSDDDND-----ETLWVE-----ATAEQVLKQPGVSVI 54
QY 65 YEEQEQEATGSRQADVAQAQLIPSLGV-----SSGTTSNFGQT-----MHGRQVQFLNGV 116
Db 55 TSEDI-KXTPPVNDLSDIIRKM-PGVNLTGNSAGSTRGNRRQIDIRGMGPENTLILIDGV 112
QY 117 PLT-----GSRDISRQLNSINPQVARIIVLSG-ATSIYSGATGGLINIVTKSD 165
Db 113 PVTSRNVSVRWGERDTRGDTNWVPPVPEQVERIEVIRPAAARYGSGAAGGVNIIITRP 172
QY 166 LEEQFETRIGVHSGKLSSEGIGYQVGQSVAGVSENGVNLARLDVY--RTTGGAFDANGK 224
Db 173 TNDMHGSLSYTNQSPESSEGAATRRANFSLGPLAGDALTTRLYGNLKNKTADSDIN-- 230
QY 225 RIAPPAOTDK-----QDSKLSVNTVNDWOLDKQKQINLALTHYNDKQDTPADYDYGRL 280
Db 231 ----SPVGTKNAGHEGVNRKQDINGVISWKLNPQILDPEAGY--SRQGNIYAGDTQNS 284
QY 281 AVLFGCKPSLNAIKGLSLSEQPKTKSTFNINY--HHDDLW 319
Db 285 SSATVE-----SLAKSGKETNRLRYQNLGTHNGIW 315

```

RESULT 23  
Q8RMB4 PRELIMINARY; PRT; 726 AA.  
AC Q8RMB4  
DT 01-JUN-2002 (TREMblrel. 21, Created)  
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
DE TonB dependent outer membrane siderophore receptor.  
GN IRON.  
OS Salmonella cholererae-suis (Salmonella enterica).  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
NCBI\_TaxId=591;  
OK NCB1\_TaxId=591;  
RN  
RP SEQUENCE FROM N.A.  
RA Wu W.-S., Chang C.-F., Chang Y.-F.;  
RT "Cloning and characterization of iron-regulated genes in Salmonella  
RT cholererae-suis";  
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.  
RW EMBL; AY029471; AAK3133.1; -.  
KW Receptor.  
SQ SEQUENCE 726 AA; 79369 MW; 5C9CD20C679EB222 CRC64;  
Query Match 9.5%; Score 195; DB 2; Length 726;  
Best Local Similarity 23.6%; Pred. No. 0.00014;  
Matches 81; Conservative 63; Mismatches 135; Indels 64; Gaps 16;  
QY 7 FQWLSPLISVAVTQOLYA--QPNESLPVLEPVIITIDKSGMALANRITOMPHTTKYI 64  
DB 8 FVWL-ITVSTGVNSPLSAESTDNGETMVE-----STAEQVLKQPGVSII 55  
QY 65 YEEQIOEQATGSRQLADVAQAOLIPSLGVSSGTTSNFGQTMHGROVP-----LLN 114  
DB 56 TRDDIQKNP-VNDLADIRKM-PGVNLTGNSAS--GTRGNRQIDIRGMGPENTLVLD 111  
QY 115 GVPILT-----GSRDISRQLNSINPQVARIIVLSG-ATSIYSGATGELINIYTK 163  
DB 112 GVPVTSRNSVRYSWGRERDTRGDTNWVPEMVERIEVIRGPAARVAGSGAAGGVNIITK 171  
QY 164 SDEEBOFETRIGVHGSKLSSEGIQVQGVASVSENGVNLARLDVY-RTTGAFFDAN 222  
DB 172 RPTDWHGSLSLYTNQPESSKEGDTTRGNFSLSGPLAGDTLTMRLVGNLRTDADSWDIN 231  
QY 223 G----KRIAPERQDTKQSKSLSVNTVNDQDKONINLALTHYNDKODTDYADYGN 278  
DB 232 SAGCTGNAAGRGVTK-----DINSVFSKMTPTQOIIDFAGY--SRGNIYAGDTQN 283  
QY 279 RLAVLFGKPSLNAIKGLSISEQPKTKSTFNINY--HHDDLW 319  
DB 284 STS-----NAVY-KSLAOSGRETNRLYRONYGLTHNGIW 316  
RESULT 24  
Q8ZMNO PRELIMINARY; PRT; 726 AA.  
AC Q8ZMNO  
DT 01-MAR-2002 (TREMblrel. 20, Created)  
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
DE TonB-dependent siderophore receptor protein.  
GN IRON OR STM2777.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
NCBI\_TaxId=602;  
OK NCB1\_TaxId=602;  
RN  
RP SEQUENCE FROM N.A.  
RC STRAIN=LT2 / SGGC1412 / ATCC 700720;  
RA MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Lattelle P.,  
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,

RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
RT LT2.";  
RL Nature 413:852-856(2001).  
DR EMBL; AE008826; AAL21663.1; -.  
DR InterPro; IPR000568; ATPrxnc\_Nsub.  
DR InterPro; IPR000531; TonB\_boxC.  
DR Pfam; PF00593; TonB\_boxC; 1.  
DR PROSITE; PS00449; ATPase\_A; 1.  
DR PROSITE; PS01156; TonB\_DEPENDENT\_REC\_2; UNKNOWN\_1.  
KW Receptor; Complete Proteome.  
SQ SEQUENCE 726 AA; 79401 MW; C47CB4075AF802A1 CRC64;  
Query Match 9.5%; Score 194; DB 16; Length 726;  
Best Local Similarity 23.6%; Pred. No. 0.00017;  
Matches 81; Conservative 63; Mismatches 135; Indels 64; Gaps 16;  
QY 7 FQWLSPLISVAVTQOLYA--QPNESLPVLEPVIITIDKSGMALANRITOMPHTTKYI 64  
DB 8 FVWL-ITVSTGVNSPLSAESTDNGETMVE-----STAEQVLKQPGVSII 55  
QY 65 YEEQIOEQATGSRQLADVAQAOLIPSLGVSSGTTSNFGQTMHGROVP-----LLN 114  
DB 56 TRDDIQKNP-VNDLADIRKM-PGVNLTGNSAS--GTRGNRQIDIRGMGPENTLVLD 111  
QY 115 GVPILT-----GSRDISRQLNSINPQVARIIVLSG-ATSIYSGATGELINIYTK 163  
DB 112 GVPVTSRNSVRYSWGRERDTRGDTNWVPEMVERIEVIRGPAARVAGSGAAGGVNIITK 171  
QY 164 SDEEBOFETRIGVHGSKLSSEGIQVQGVASVSENGVNLARLDVY-RTTGAFFDAN 222  
DB 172 RPTDWHGSLSLYTNQPESSKEGDTTRGNFSLSGPLAGDTLTMRLVGNLRTDADSWDIN 231  
QY 223 G----KRIAPERQDTKQSKSLSVNTVNDQDKONINLALTHYNDKODTDYADYGN 278  
DB 232 SAGCTGNAAGRGVTK-----DINSVFSKMTPTQOIIDFAGY--SRGNIYAGDTQN 283  
QY 279 RLAVLFGKPSLNAIKGLSISEQPKTKSTFNINY--HHDDLW 319  
DB 284 STS-----NAVY-KSLAOSGRETNRLYRONYGLTHNGIW 316  
RESULT 25  
Q93SS7 PRELIMINARY; PRT; 688 AA.  
AC Q93SS7  
DT 01-DEC-2001 (TREMblrel. 19, Created)  
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)  
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)  
DE Huga.  
GN HUGA.  
OS Plesiomonas shigelloides (Aeromonas shigelloides).  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Plesiomonas.  
NCBI\_TaxId=703;  
RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21189295; PubMed=11292789;  
RA Henderson D.P., Wyckoff E.B., Rashidi C.E., Verlei H., Oldham A.L.;  
RT "Characterization of the Plesiomonas shigelloides genes encoding the  
RT Heme Iron Utilization System.";  
RL J. Bacteriol. 183:2715-2723(2001).  
RW EMBL; AY008342; AAG21395.1; -.  
RW InterPro; IPR000531; TonB\_boxC.  
RW Pfam; PF00593; TonB\_boxC; 1.  
SQ SEQUENCE 688 AA; 76073 MW; C70E7088F1C6CCCC CRC64;  
Query Match 9.3%; Score 189.5; DB 2; Length 688;  
Best Local Similarity 28.5%; Pred. No. 0.00031;  
Matches 79; Conservative 53; Mismatches 96; Indels 49; Gaps 17;  
QY 10 LSLPLFL-SVAVTQOLYAQPNESLPVLEPVIITIDKSGMALANRITOMPHTTKYIYEEQ 68  
DB 11 LSLPLFL-SVAVTQOLYAQPNESLPVLEPVIITIDKSGMALANRITOMPHTTKYIYEEQ 68

```
Db 32 LSLILNSVAVSPAANAFAE--PQTQLNEVLVTATRENKAL-----SQETRSVAVVTOAQ 85
Qy 69 IQEQATGSRQLADVMAQLIPSLGVSSGTTSNFG-----QTMHGRQVQLNGV-PLTGRSD 123
Db 86 IQQQQPAS--VAEAL-KYIPNDVSGPRANAQSPVRLSGNRIILQVVVDGVRQNTTSGH 142
Qy 124 ISRLNSINPNQVARIEVLISGAT-SIYSGATGGLI--NIVTKSDLEEEQFETRIGVHGS 180
Db 143 V--RLTSWDFPEMLSSIEVTKGPTSSUWGSGALGVAAQQTISAPDN-----LHG- 189
Qy 181 KLSSEGIGYQVGSVAGVSE---NGNVLARL--DVDYRTTGGAFDANGKRIAPETPAQTD 234
Db 190 ---DQSVGGYLKQYASANDETKTSVAVGLLGDSDVDFLLNGYADANNLRLNGEHLTD 246
Qy 235 K--QDSKLSLVNTVDWQDDKQINLAL-----THYN 265
Db 247 SAYRNOGGLA---KFGWQADDAQRLELSVRQAETHQN 280

RESULT 26
QBYRY2
ID Q8YRY2 PRELIMINARY; PRT; 698 AA.
AC Q8YRY2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Hypothetical protein A113310.
GN AL13310.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpō S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003592; BAB75009.1; -
DR InterPro; IPR00531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 698 AA; 76888 MW; 58D36565E58145D1 CRC64;

Query Match 9.2%; Score 187.5; DB 16; Length 698;
Best Local Similarity 25.0%; Pred. No. 0.00043;
Matches 91; Conservative 61; Mismatches 141; Indels 71; Gaps 21;

Qy 6 YFOWLSPLLSVAVVTOQL-----YAQPNEISLPTVE---LEPVITIDKSGMALANRIT 55
Db 39 YLSEITLPTTSABLLTQSTPGDLNPEGPSEQ-PEIETSSDDADITIE---ATAEPET 93
Qy 56 QNPHT-TKVIYIEQIQEQATGSRQLADVMAQLIPSLGVSS---GTTSNFGQTMHGRQVQ- 110
Db 94 LPASTPTTYIDQEIQKQ--GATSVADVLRKM-PCGAINDVGHGADIHTGTYYRGASINQ 150
Qy 111 --FLNGVPLTGSRD1-----SRQLNSINPNQVARIEVLISGATS-IYSGATGGLINI 162
Db 151 SVFLNGRI--NNDVNTHTGATDLSIPVESIERVELSGVTSALYGSAGFGVNNIT 208
Qy 163 KSDLBEEQFETRIGVHGSKLSSEGIGYQVGSVAGVSENGNVLARDVD--YRTTGGAPD 220
Db 209 KKGYPQPKLTSSLEFGLSLNNQOFSY---SGAVGAATVNFSPERYFVNDNRVVPVGA-- 263
Qy 221 ANGKRIAPETPAQTDKQ-----DSKLSLVNTVDWQDDKQINLALTHYNDKQDQTDY-- 272
Db 264 -----ANRDSQGFLSNADTSTSYFGNIGLDQDRLSLSDITKLSRRGLVYFG 313
Qy 273 APDYGNNLAVLFGKEKPSLNAIKGLSSEQ-----PKTKSTFTFNINVHDDLMGNTINT-N 326
Db 314 FLQDRDL-----DHDGLNI--GLSWKTRLGNNGNSNLTTTGYNQNYSTYGTPTVPAGR 366
```

```
Qy 327 AYR 330
Db 367 BFR 370

RESULT 27
QBYTPO
ID Q8YTP0 PRELIMINARY; PRT; 867 AA.
AC Q8YTP0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ferrichrome-iron receptor.
GN AL12674.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpō S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003590; BAB74373.1; -
DR InterPro; IPR00531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 867 AA; 95468 MW; 5D21556D8E781FEA CRC64;

Query Match 9.1%; Score 187; DB 16; Length 867;
Best Local Similarity 23.9%; Pred. No. 0.00063;
Matches 87; Conservative 52; Mismatches 131; Indels 94; Gaps 15;

Qy 10 LSLPLLSVAVVTOQLYAQPNEISLPTVEL-----EPWVITI--- 43
Db 130 VSLPVTTSQPTTQTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 189
Qy 44 -----DKSGMALANRITQMPHTTKVIYIEQIQEQATGSRQLADVMAQLIPSLGV 93
Db 130 QRPPTYQVENS SVGTRTDTPLIDVPQAIQVPEVIKQD--GTRSLGEVLKN---TSSAS 244
Qy 94 SGTTSN---FCQTMHGRQVQLNGVPLTGSRDISRQLNSINPNQVARIEVLISGATS- 148
Db 245 SGRTSSQAPALTPTVIRGPESRNLRLR-----NGLRDSDLRFQSEIAN-VERVEVLKGPASVL 299
Qy 149 YSGATGGLINIIVTKSDLEEEQFETRIGVHGSKLSSEGIGYQVQVQ-----SVAGVSE 200
Db 300 FQCGDLGGVNLVTKQPLNTPY-----SIGYQVGFGLHRTIDFSGGLD 345
Qy 201 NGNVLARDVDYRTTGGAFDANGKRIAPETPAQTDKQDSKLSLVNTVDWQDDKQINL- 259
Db 346 KGVAVRLNAAVQTA-----ESFKDFENSESFPIAPV--RLIGNENTNLT 389
Qy 260 ALTHYNDKQDQTDYAPDYGNNLAVLFGKEKPSLNAIKGL---SLSE-OPKTKSTFTFNIN 315
Db 330 ASIEYLVKRSFETAPDLPASGTVISNPNGRVSRNTRNLPESLSESLVTRLGQLDHL 449
Qy 316 DDLW 319
Db 450 DNDW 453

RESULT 28
O86903 PRELIMINARY; PRT; 747 AA.
ID O86903
AC O86903;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
```

DT 01-NOV-1998 (TRENBLREL. 08, last sequence update)  
 DT 01-DEC-2001 (TRENBLREL. 19, last annotation update)  
 DE Ferrichrome-iron receptor.  
 GN PHU.  
 OS Salmonella paratyphi-a.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Salmonella.  
 OX NCBI\_TaxId=54388;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SL369;  
 RX MEDLINE=98348446; PubMed=9683481;  
 RA Kallmann H., Herrmann C., Wolfe H., Braun V.;  
 RT "Identification of a new site for ferrichrome transport by comparison  
 of the PhnK proteins of Escherichia coli, Salmonella paratyphi B,  
 RL Salmonella typhimurium, and Pantoea agglomerans."; J.  
 Bacteriol. 180:3845-3852(1998).  
 DR EMBL; Y14067; CAA74395.1; -.  
 DR HSP; P06971; 18Y5.  
 DR InterPro; IPR000531; TonB boxC.  
 DR Pfam; PF00593; TonB\_boxC.1.  
 DR PROSITE; PS00430; TONB DEPENDENT\_RBC\_1; UNKNOWN\_1.  
 DR PROSITE; PS01156; TONB DEPENDENT\_RBC\_2; UNKNOWN\_1.  
 KW Receptor.  
 SQ SEQUENCE 747 AA; 82292 MW; 3F05A4386391A30C CRC64;

Query Match 9.1%; Score 186; DB 2; Length 747;  
 Best Local Similarity 24.2%; Pred. No. 0.00059;  
 Matches 98; Conservative 50; Mismatches 139; Indels 118; Gaps 21;

QY 25 ACPNBSLPVLEPEVITIDKSGALANRITOMP--HTTKVIEEOIOEQATG-----S 76  
 DB 7 ACPNBSLPVLEPEVITIDKSGALANRITOMP--HTTKVIEEOIOEQATG-----S 76  
 QY 77 RQIA-----DVAQQLP-----SLGVSST--TSNF-----100  
 DB 64 ROSATATKTDTPLOKVPQISVVTAEEMALHQPYSVEALSTYPPGVAVGTGASNTYDY 123  
 QY 101 ---GQTHGRQVQFLNGVPLTGS--RDISRLNSINPNQVARIIVLSGATSI--YSGAT 154  
 DB 124 IIRGPAAGGSONNYINGLMGOQNFYND-----VIDPVLLEAEVWRGVSVLYGKSS 178  
 QY 155 GGLINIVTK---SDLEEQFETRIGVHSGKLSSEGIQVGVSVGSENGVTLARLDV 210  
 DB 179 GGLINIVTK---SDLEEQFETRIGVHSGKLSSEGIQVGVSVGSENGVTLARLDV 210  
 QY 211 DVRTTGAPDANGRIAPAPQTDKQDSKLSVNTVNDWQLDKQINIALTHYNDKQDT 270  
 DB 225 -YRLTGLASSAN-----AQODRAEEGRYAIAPAFWRPDDKTPTF--LSYFQNEP 274  
 QY 271 DVAPDGNRLAVLFGKPSINAKGLSLSEQPTTSTFNINHHDDLMGNTINTNAYTR 330  
 DB 275 GY-----YGMVPEKGYEPL--PNGRLPLDFE-----EGAKNTYSR 310  
 QY 331 REKGRFYPVAPPSIAKALPILOSMNLPASATLDAYTKAPQARAYG 375  
 DB 311 NEMWGVSPDHEFN--DITVRQNLRY-----AENKVSQNSVYG 347

RESULT 29  
 OBYTT7 PRELIMINARY; PRT; 872 AA.  
 DT 01-MAR-2002 (TRENBLREL. 20, Created)  
 DT 01-MAR-2002 (TRENBLREL. 20, last sequence update)  
 DT 01-JUN-2002 (TRENBLREL. 21, last annotation update)  
 DE Ferrichrome-iron receptor.  
 GN ALR2626.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OX NCBI\_TaxId=103690;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=21595285; PubMed=11759840;  
 RA Kaneko T., Nakamura Y., Wolk C.P., Kurlitz T., Sasamoto S.,  
 RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsuno M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yanada M.,  
 RA Yasuda M., Tabata S.;  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 RT cyanobacterium Anabaena sp. strain PCC 7120.";  
 RL DNA Res. 8:205-213(2001).  
 DR EMBL; AP003590; BAB7435.1; -.  
 DR InterPro; IPR001484; Pyrokinin.  
 DR InterPro; IPR000531; TonB boxC.  
 DR Pfam; PF00593; TonB\_boxC.1.  
 DR PROSITE; PS00539; PYROKININ; UNKNOWN\_1.  
 KW Receptor; Complete proteome.  
 SQ SEQUENCE 872 AA; 95972 MW; 6485E569BEA3E61 CRC64;

Query Match 8.9%; Score 182; DB 16; Length 872;  
 Best Local Similarity 22.4%; Pred. No. 0.0014;  
 Matches 78; Conservative 62; Mismatches 130; Indels 78; Gaps 16;

QY 11 SUPLSVAVTQQLVQNPESLPVLEP-----VVITIDKSGMALNR---53  
 DB 163 AVPTVTTQTPQTPVPEPQSTETQBPQSGAPEAPLHLLVYGQNRVYVNPASTG 222  
 QY 54 -----ITOMPRTKVIYEEQIOEQATGSRQADVMAOL--IPSLGVSS-----GTTSN 99  
 DB 223 TWTDLIRIDIPITQIVVEQVAKQDV--TRLRDLNINIGVVGQGGSTSDQIGINGF 280  
 QY 100 FQQTHGRQVQFLNGVPLTGSRDISRLNSINPNQVARIIVLSGATSI--YSGATGLI 158  
 DB 281 FQDGFPGGSI--LVQDFK--DGRGIRETAN-----VERIEVLKGPASVLYGQVPGGV 331  
 QY 159 NIVTSKDEEOPFETRIGVHSGKLSSEGIQVGVSVGSENGVTLARLDVPRITCG 217  
 DB 332 NLVTQPLDPYVNAELVS--GSFST-----FRPSIDISQPLNSDKTLRLNSVETSDG 385  
 QY 218 APDANGRIAPAPQTDKQDSKLSVNTVNDWQLDKQINIALTHYNDKQDT--YAP 274  
 DB 386 FRDFN-----QDVGRFPISPLKWKIEIKATNLTQFPLYNDERFPDGFALF 432  
 QY 275 DYG---NRLAVLFGKPSINAKGLSLSEQPTTSTFNINHHDDLM 319  
 DB 433 GGIITDPLERFFGEPDDVRYKVEIGLS-----YLEHNFNDNW 471

RESULT 30  
 OBYMK9 PRELIMINARY; PRT; 858 AA.  
 DT 01-MAR-2002 (TRENBLREL. 20, Created)  
 DT 01-MAR-2002 (TRENBLREL. 20, last sequence update)  
 DT 01-JUN-2002 (TRENBLREL. 21, last annotation update)  
 DE Ferrichrome iron receptor.  
 GN ALA4924.  
 OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OX NCBI\_TaxId=103690;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21595285; PubMed=11759840;  
 RA Kaneko T., Nakamura Y., Wolk C.P., Kurlitz T., Sasamoto S.,  
 RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsuno M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yanada M.,  
 RA Yasuda M., Tabata S.;  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 RT cyanobacterium Anabaena sp. strain PCC 7120.";  
 RL DNA Res. 8:205-213(2001).  
 DR EMBL; AP003598; BAB76623.1; -.  
 DR InterPro; IPR000531; TonB\_boxC.  
 DR Pfam; PF00593; TonB\_boxC.1.  
 KW Receptor; Complete proteome.

```
SQ SEQUENCE 858 AA; 94913 MW; 8F227AFCEA3B7918 CRC64;
Query Match 8.9%; Score 181.5; DB 16; Length 858;
Best Local Similarity 24.5%; Pred. No. 0.0014;
Matches 91; Conservative 51; Mismatches 151; Indels 79; Gaps 17;

QY 18 AVTQQLYQAPNESLPTV---ELEPVVITIDKSGMALANRITQMPHTTKYVIEEQIQEQAT 74
DB 177 AIPQAPQAPDEPIELVVTGEOQOYRPAASVGTTRDTPLRDIPOSIQVVPQVLOEQR- 235

QY 75 GSRQADVMAQLI---PSLGVSSGTTSNFGQTMHGQV---QFLNGVP---LTGSRDTS 125
DB 236 -ATRLGDLARNVSGVNPTRG-SGDRADSP--TIRGEIFSGVNLNGLDRLTETREDI- 290

QY 126 ROLNSINPNQVARIELVSGATSI-YGSGATGGLINIVTKSDLEEEQFETRIGVHGSKLS 184
DB 291 -----YNERVEVLKGPASVLYGLGNPGGTNVITKQPLANPPYIEATVGNVDLYR 342

QY 185 EGIGYQVQSGVAGVSENGVNLARLDVDYRTTGGAPANGK---IAP----- 228
DB 343 GGIDFS-----GFLNDSKTLRLNLAAYQNSGSYIDFVGNRSFFIAPVISAALGNKNTLT 397

QY 229 -EPAQTDKO-DKSLSVNTNVQWL---DKONINLALTHYNDKQDTDY-APDYGRLAVL 283
DB 398 FEGEYSQKIDSTRTVVVLPAVGTVPGPGRIPRNRVTVIEPEGDTQIETRLGYRLEHR 457

QY 284 PQEKPSLNAIKGLSLSEQPKTKSTFNINYHHD-----DLWGNTINTNAYRRE 332
DB 458 FSENWSL-----RNDPRVTFEHNADNNQAFLLGLDADNRNTANRSTYSSES 502

QY 333 KGRFYFVAPFS 344
DB 503 DSNINYNLTTDIS 514

RESULT 31
QX9BV1 PRELIMINARY; PRT; 708 AA.
AC Q9XBV1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TonB-linked receptor Tlr.
GN TLR.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; CF6 group; Bacteroidetes; Bacteroidales; Porphyromonadaceae;
OC Porphyromonas.
OX NCBI_TaxID=837;
RN [1]_TaxID=837;
RP SEQUENCE FROM N.A.
RC STRAIN=WS0;
RX MEDLINE=21069349; PubMed=11154437;
RA Slakeski N., Dashper S.G., Cook P., Poon C., Moore C., Reynolds E.C.;
RT "A Porphyromonas gingivalis genetic locus encoding a heme transport
RT system.";
RL Oral Microbiol. Immunol. 15:388-392(2000).
DR EMBL; AF155223; AAD37808.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Receptor.
SQ SEQUENCE 708 AA; 78906 MW; 1B19E61A2CB9B311 CRC64;

Query Match 8.8%; Score 180.5; DB 2; Length 708;
Best Local Similarity 23.4%; Pred. No. 0.0013;
Matches 94; Conservative 68; Mismatches 131; Indels 109; Gaps 21;

QY 14 LLSVAVTQQLYQAO----PNESLPTVELEPVVITIDKSGMALANRITQMPHTTKYVIEEQI 69
DB 8 LLSIGISQAFAKTDNVPDTSRVNLNLTQTVYSTRVAPL-----KKIPAKMELISSRNI 63

QY 70 QEOATGSRQADVMAQLIPSLGVSS--GTTSNF---GQTMHGROVQFLNGVPLTGSRI 124
DB 64 KQ--SGFNNTUIL-KTQSSLDVIQYFGSSNIGIRGPKSKYTVLVNGIP-AGTDNI 119
```

```
QY 125 SRQLNSINPNQVARIELVSGA-TSIYSGATGGLINIVTKSDLEEEQFETRIGVHGSKLS 183
DB 120 S-----TLNITSIEQTEILKGPFSIYGTNMGVNVNITTHSKDK-----IHGN-VS 166

QY 184 SEGIGYQVQSGVAGVSENGN-----VLARLDVDYRTTGGAFDANGKRIAPEPA 231
DB 167 LFGGSY---QTMAGSFNLGGRFEDIFSFLSLGLDKQNKDYKTGSNNFLSLSLK-----BEA 220

QY 232 QTDKQDSKS-----LSVNTNVQWLDDKQKONINLALTHYNDKQDTDYAPD 275
DB 221 IVDVNAATNKKMKGSDYTVATGRRLRFGDPTPEWSLNLVQNVFLG-----DALPV 270

QY 276 YGNRLAVLFGKPKSLNAIKGLSLSEQPKTKSTFNINYHDDLMGNTINTNAYRREKGR 335
DB 271 GGSINGVYGESKKNLN-----RSSTSPPELLKHG---CHTLQFSPYFNIEKSE 315

QY 336 FYFVAPFSIAKALPILQSMNLPSATLDAYTKAQAARAYGV 377
DB 316 NYNNADPTGF-----INYKS---DYTT-----YGAL 338

RESULT 32
QX9UZ6 PRELIMINARY; PRT; 853 AA.
AC Q8YUZ6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ferrichrome-iron receptor.
GN ALR2185.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003588; BAB73884.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 853 AA; 94548 MW; 67EA77C09E049437 CRC64;

Query Match 8.8%; Score 180; DB 16; Length 853;
Best Local Similarity 23.2%; Pred. No. 0.0018;
Matches 82; Conservative 66; Mismatches 118; Indels 88; Gaps 19;

QY 11 SLPLSVAVTQQLYQAPNESLPTVELEPV--VITIDKSGMALANRIT-----QMPH 59
DB 172 SLP-----PQPETQSQSPSPAAB-EPIELVVTGKDGQVQVPNTATVTRTNTPIIDIPQ 224

QY 60 TTKYVIEEQIQEQATGSRQADVMAQLIPSLGVSSGTTSNFGQ--TMHGROVQFLNGVP 117
DB 225 SIQVVPVQVLESDQ-----QITRVDALRNVPGLSGTNAFTGNQITING-----FSTSLP 275

QY 118 LTGSRD---ISRLNSINPNQVARIELVSGATSI-YGSGATGGLINIVTKSDLEEEQFET 173
DB 276 IL--RDGFRFIEYENSPQTSNLEKIEVLKGPASVQYQGLDPGGVNLVTKPLSEPPFEI 333

QY 174 RIGVHGSKLSSEGIGYQVQSGVAG--VSENGVNLARLDVDYRTTGGAFDANGKRIAPEPAQ 232
DB 334 Q-----AQFGSVGL--IRPSFDVSGPLTDDKGLYRLNATYQREEGFRDN----- 377

QY 233 TDKQDSKLSVNTNVQWLDDKQKONINLALTHYNDKQDTDYAPDYGRLAVLFGKPKSLNA 292
```

```

Db 378 ---TETTERFIADSLTWKISDRTNVPSLEYLSTRPEDTS-----LVAFER----- 421
Qy 293 INGLSISEOPKT-----TKS---TFNINHHDDLGMNTINNAVYRRE 332
Db 422 ----SVADVPYSVFPDPPDFITKSPSTAYNLEHPSDNW--TLANSFRTYLOQ 469

RESULT 33
Q9KIB0 PRELIMINARY; PRT; 757 AA.
AC Q9KIB0;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE Hypochemical conb-linked outer membrane receptor PG13.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; CFB group; Bacteroidetes; Bacteroidales; Porphyromonadaceae;
OC Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H50;
RA Rose B.C.;
RT "Genomic analysis of Porphyromonas gingivalis for vaccine discovery.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF237558; AAF81416.1; -.
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC; 1.
KW Receptor.
SQ SEQUENCE 757 AA; 84517 MW; 224B6D5264F9D62 CRC64;

Query Match 8.7%; Score 178.5; DB 2; Length 757;
Best Local Similarity 25.1%; Pred. No. 0.0019;
Matches 83; Conservative 56; Mismatches 147; Indels 45; Gaps 16;

Qy 31 LPTVELPEVITIDKSGMALANRITQMPHTTKYIEQIOEQATGSRQLADVNAQ-LIPS 89
Db 106 LRTNNLEEVVVT-----GTGRYRLVAPVATEVLTAKD---ASFSAPTSEALLQGLSPS 158
Qy 90 LGVSSGTTSNFGQT--MHGRQVQFLNGVPLTGSRDISRLQNSINPQVARIIVLSGA-T 146
Db 159 FDEGPNLMGSEFMQNLNGLSKSYIILIDGKRVYDVGQADLSRISPDQIRIELIVGASS 218
Qy 147 SIYSGATGGLINIVTKSDLEEOFT--RIGVHGSKLSESGIGYGVGSVAGVSENGV 204
Db 219 SLTGSDAIACVAVITKQNTNRLSAVTSKRISKNDKQVTSIDINIGK---FSSNTNY 274
Qy 205 LARLDVYRTTG--GAFDANGKRIAPAPA--QTDKQDSKSLVNTVNDQLDKQINIL 259
Db 275 FF-----YHTDGMQNSPELKKKKGSGEPVLEETYYKTFPA-QENGCVSSLSYATNNL 328
Qy 260 ALT--HYNDKQDTYAPDYGNRLAVLFGKPSLAIKGI-----SLSEOPKTKTS 307
Db 329 SFGSNVQYNNRQ--IFPTPSEKKAYMDYR-ALTASLGTNYLPFNGLHTLSFDAYVDRF 385
Qy 308 TFNINHHDDLGMNTINNAVYRREKGRFYP 338
Db 386 RFGYLTHDKDSSESLINNQG--QTEOPTFPF 414

RESULT 34
Q8YVZ8 PRELIMINARY; PRT; 802 AA.
AC Q8YVZ8;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

```

```

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Outer membrane heme receptor.
GN ALR2153.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,
RA Matanabe A., Iriuchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muzaki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takezawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003588; BAB73852.1; -.
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 802 AA; 89572 MW; 513804669981AC30 CRC64;

Query Match 8.7%; Score 178.5; DB 16; Length 802;
Best Local Similarity 25.3%; Pred. No. 0.002;
Matches 94; Conservative 54; Mismatches 162; Indels 61; Gaps 17;

Qy 35 ELEPVITIDKSGMALANRITQMPHTTKYIEQIOEQATGSRQLADVNAQLPSLGVS- 93
Db 95 ENEEADIEITVGTTRPSLSDSPGTTIVIDABEDIDNQI---QNLDDLRYEP--GVST 149
Qy 94 SGTTSNFG-----QMHGRQVQFLNGVPLTGSRD-I-SRP--NSINPQVARIIVLSG 144
Db 150 SGDAARYGQDNRINRIGDENRVLQVDGVRLPDSFPGSTQLGKRVYIDETLRVPIING 209
Qy 145 -ATSIYSGATGGLINIVTK--SDLEEOFTFTRIGVHGSKLSESGIGYGVGSVAGVSEN 201
Db 210 SASITLYGSDAIGGVVFTFKDPSDYLINEGDD--GVFSNKFVYDSSANRIGWTTIAGRL 267
Qy 202 GNVLAALDVYRTTGAFAPANGKRIAPAPAQTDKQDSKSLVNTVNDQLDKQINILAL 261
Db 268 GDVEGLL-IYTRRDGYEPOINSR-APNPQTTDAN---SWLTQLVFLGDFDQKLTG 320
Qy 262 THYNDKQDTYAPDYGNRLAVLFGKPSLAIKGLSISEOPTKSTFNINHHDDLGMN 321
Db 321 EFINRTTIDNVLTSG--INFQ-----VRTSLTITDEIKKRYLSTEH-----N 364
Qy 322 TINTNAVYRREKGRFYPFVAPPSIAKALPILQSMULPSATLDAYTKAPQARAY----- 374
Db 365 NPDNSLFPQVLASQIY-----YQATSTBSNELRRATAPITTGAVNRFRPRRSIYQ 417
Qy 375 ----GYLQSES 381
Db 418 NTFGGDIQLDES 428

RESULT 35
Q8RFD8 PRELIMINARY; PRT; 715 AA.
AC Q8RFD8;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hemin receptor.
GN FN0768.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
OX NCBI_TaxID=76856;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25586;
RX MEDLINE=2186394; PubMed=11889109;
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,

```







```
QY 233 TDKQSKSLSVNTVDWQDDKQ----NINLALHYNDKQ 268
Db 231 SPTK-ANNFNIGTRISYLANDYNTFFIDFSRNHYDNKQ 269

RESULT 40
Q9PMG3 PRELIMINARY; PRT; 696 AA.
AC Q9PMG3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Putative iron uptake protein.
GN CFPA OR CJO755.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCCT 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
DR EMBL; AL139076; CAB73021.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Complete proteome.
SQ SEQUENCE 696 AA; 77554 MW; 31E09CB15187388C CRC64;

Query Match 8.2%; Score 168.5; DB 16; Length 696;
Best Local Similarity 25.0%; Pred. No. 0.0075;
Matches 70; Conservative 57; Mismatches 104; Indels 49; Gaps 15;

QY 14 LLSVAVTOOLYQAQNESLPTVELEPVVITIDKSGMALANRITQMPHTTIVYEEQIQEA 73
Db 14 LSSNAISQ-----NVELDSSIV---SASGFTODIKEAPATINVTIKELQSKP 58

QY 74 TGSROLADVMAQLIP--SLGVSSGTTSNFGQMHG--RQVQLNG-----VPLTGS 121
Db 59 --YRDVAEAIAD--IPGVDLVYASKGKTGTYNITMRGITGYTLVLIDGRROGIGGEVGPNGF 115

QY 122 RDISRLNSINP--NOVARIEVLGATS--LYSGGATGLINIVTKSLDLEEEQFETRIGVH 178
Db 116 NEIS---NSFLPPIISIERIEVKGPMSTLYGSEALGGVNNITTK--VSDRWETSLSLD 170

QY 179 GSKLSEGGIGYQVGOSVAG-----VSENGVNLARLDVYRTTGA--FDANGKRIAPEPAQ 232
Db 171 ALLNENKONGNTYGTSIYSSGPLMDKLGTLRFREFYQQSNVEFTNGSGQRVQGDQAQ 230

QY 233 TDKQSKSLSVNTVDWQDDKQ----NINLALHYNDKQ 268
Db 231 SPTK-ANNFNIGTRISYLANDYNTFFIDFSRNHYDNKQ 269

RESULT 41
Q9PMG1 PRELIMINARY; PRT; 709 AA.
AC Q9PMG1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Haemin uptake system outer membrane receptor.
GN CHUA OR CJI614.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
```

```
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCCT 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
DR EMBL; AL139079; CAB73602.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Complete proteome.
SQ SEQUENCE 709 AA; 80940 MW; E4F63AB80EF32741 CRC64;

Query Match 8.2%; Score 168.5; DB 16; Length 709;
Best Local Similarity 21.8%; Pred. No. 0.0077;
Matches 74; Conservative 65; Mismatches 116; Indels 85; Gaps 15;

QY 60 TTKVIYEQIQEQATGSRQLADVMAQLIPSL-----GVSSGTTSNFG 101
Db 26 SNKAINLOKVVSTTFEQDADSNLRNVISIEGKDLQNKGYVSLAQALERISSISFVNFG 85

QY 102 Q----TMHGR-----QVQFLNGVP---LTGSRDISRLNSINPNOVARIEVLGATS 148
Db 86 LGRNIDMGQGNKSNIAVKVMIDGHAINVLNDSHCVT--PLDSINLDNVERIEIIPGGGSV 144

QY 149 -YGSATCGGLINIVTKSLDLEEEQFETRIGVHGSKLSSSEGIYQVGQSVAQVSE--NGNVLA 206
Db 145 LYSGTGRGGVINITK---KQSDAFAINLKSSAYDHGGLGNLG--INGAKQINENLAF 199

QY 207 RLVDVYRTTGGAFDANGKRIAPEPAQTDKQSKSLSVNTVDWQDDKQKQINLALHYND 266
Db 200 SFDIQ-----SFNLDGQY-----EGYNEKGYFINTKYTIDINDNSDLTGYNYPKS 245

QY 267 KQDTPADYDGNRLAVLFGKPSLNAIKGLSISEQPKTKS-----TFNINVHH 315
Db 246 KNTS-----SGVLTAKAQSDPTQKGSNDNITQINRPEISLDYHYFF 287

QY 316 DLLMGNTINTNAYRREKGRFPYFPVAPPS--IAKALPILOS 354
Db 288 DDIW--EFNLEAFQWQKINLYLKDVSTMSYMNMLPVPYQN 325

RESULT 42
Q8YTW7 PRELIMINARY; PRT; 820 AA.
AC Q8YTW7;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Ferrichrome-iron receptor.
GN ALR2596.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Mitanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL; AP003590; BAB74295.1; -.
DR InterPro; IPR000531; TonB_boxC.
```

DR Pfam: PF00593; TonB\_boxC.1.  
 KW Receptor; Complete proteome.  
 SQ SEQUENCE 820 AA; 91035 MW; BBDOB2486468451 CRC64;

Query Match 8.2%; Score 168; DB 16; Length 820;  
 Best Local Similarity 23.2%; Pred. No. 0.01; Mismatches 124; Indels 82; Gaps 19;  
 Matches 79; Conservative 56;

10 LSLPLSLVAATQOLVAQPNESLPTVEL-----EPVVTI-----DKSGMALANRITQ----- 56  
 DB 128 LILAVSTTATQAPATPTSDAPAPRIAGQDDPIELVVGEDRRKVRPAASATKTDTP 167  
 QY 57 ---MPTTKYIEEQIOEQATGSRQLADVMAQLISLVSSGTTSGFQTMHGRQVFL 113  
 DB 188 VRDIPESIQVPIQIILEDQKT--TRIQEVLQNV---SGVNR--QGVYGGTDAG----- 233  
 QY 114 NGVPLTGSNDISRLQSLINN-----QVARIETLSGATSI-YGSGATGGLINIVTK 163  
 DB 234 -GYRIGFPDQGNFRNGFNTDPTSLVDVTANIRIEVLKRPASVLRQAPRGGINVVTK 292  
 QY 164 SDEEEOFTRIQVHSGKLSSEGIQVQGVAG-VSENGNVLARLDVDYRTTGAFFAN 222  
 DB 293 QPRTYVAAELAVNGYAF-----YRSPDISGFLTDDSLIYRLAVAYQNGSGFPDYV 346  
 QY 223 G-KRIAPPAQTDKODSKLSLVNTVMDQLDKQNTLALTHYNDKQDTDYADYGRRLA 281  
 DB 347 FLERVFVAVP-----ITWNISDRSLTFDLEY---QDNLYLFRDG---I 384  
 QY 282 VLFGEKPS---LNAIKGLSSEQKT-TKSTFNINY--HHD 316  
 DB 385 PSIGDRPAPIPISRFVGL-----PHVYNDSTFRIGYRLHD 420

RESULT 43

Q9X7K9 PRELIMINARY; PRT; 726 AA.  
 AC Q9X7K9;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 GN Outer membrane siderophore receptor precursor.  
 OS Rhizobium leguminosarum.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 OX NCBI\_Taxid=384;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=8401 PRL1;  
 RX MEDLINE=20244627; PubMed=10784041;  
 RA Yeoman K.H., Wisniewski-Dye F., Timony C., Stevens J.B., deluca N.G.,  
 RA Downie J.A., Johnson A.W.B.,  
 RA "Analysis of the Rhizobium leguminosarum siderophore-uptake gene  
 RT fhvA differential expression in free-living bacteria and nitrogen-  
 RT fixing bacteroids and distribution of an fhvA pseudogene in different  
 RT strains."  
 RL Microbiology 146:829-837(2000).  
 DR EMBL: AJ238208; CAB41037.1; -.  
 DR HSSP: P06971; IQOQ.  
 DR InterPro: IPR000531; TonB\_boxC.  
 DR Pfam: PF00593; TonB\_boxC; 1.  
 KW Receptor; Signal.  
 FT SIGNAL 1 37 POTENTIAL.  
 SQ SEQUENCE 726 AA; 79499 MW; BDBFP2A5C5AA0408 CRC64;

Query Match 8.2%; Score 167.5; DB 2; Length 726;  
 Best Local Similarity 22.5%; Pred. No. 0.0093;  
 Matches 100; Conservative 62; Mismatches 146; Indels 137; Gaps 24;

QY 7 FQWLSPLSLVAATQOLVAQPNESLPTVELPVI-----TIDSGMALANR----- 53  
 DB 21 FVTTAIVLIGIAASPAASQATADASATA-LEPIVIGGAASDSKADRTSVAANKSSAATK 79

QY 54 ----ITOMPHTTKYIEEQIOEQATGSRQLADVMAQLISLVSSGTTSGFQTMHGRQV 109  
 DB 80 INTPLVETPRSVSVTTEKEIEOR--GAQSIIBAVRY---SAGVTTGPNG-----FDPFRD 129  
 QY 110 QPLNVPPLT---GSRDISRQ-----LNSINPNQVARIETLSGATSI-YGSGATGGLIN 159  
 DB 130 QIFIRFENITTVGDYRDSLRQPIYNGMFRTPYQIQREVIKGPVSVLYGSSPGGLVN 189  
 QY 160 IYTKSDLEEOFTRIQVHSGKLSSEGIQVQGVAGVSENGNVLARLDVD--YRTTGG 217  
 DB 190 KISKLPTEERPIHV-----GISYTKDRAQAMFDCGPISEGNDELPIYIVGL 237  
 QY 218 A-----PDANGR-IAP-----EP-----AQTDKODSKLSLVNTVMDQ-LDD 253  
 DB 238 ARHGDNPNFIADDRYFLAPSLFTWKPDGTSFTLYGLAQSDERTDA-NVGAITTVDGKILDI 296  
 QY 254 KQINLALTHYNDKQDTDYADYGRRLALFPEKPSLNAIKGLSSEQKTSTFNIN 313  
 DB 297 ROS-----DPDI--DYQYKQOQIGYQFHEPDNGLTFRQ-----NLRY 333  
 QY 314 HHDDL-----MGNTI-NTNAYYRREKGRFPVAPFSIAKALPILOSMNLPSATLD 363  
 DB 334 SHUDDLARILGVSSWTGTVAHRNA-----SSIRDEMNVPQVDNQLAKFD 378  
 QY 364 A-----YTKAPQARAYGV 376  
 DB 379 TGPLAHTMLFGLDYTNLQSNMGYGI 403

RESULT 44

Q93JP3 PRELIMINARY; PRT; 747 AA.  
 AC Q93JP3;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 GN Ferric hydroxamate uptake receptor.  
 OS Rhizobium leguminosarum (biovar viciae).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 OX NCBI\_Taxid=387;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Carter R.A.;  
 RT "Structure, function and regulation of the Rhizobium leguminosarum vbs  
 RT gene, which specify the synthesis of the siderophore vibriacin."  
 RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AJ315451; CAC48054.1; -.  
 DR InterPro: IPR000531; TonB\_boxC.  
 DR Pfam: PF00593; TonB\_boxC; 1.  
 KW Receptor.  
 SQ SEQUENCE 747 AA; 82183 MW; F6FB332B01AB4F92 CRC64;

Query Match 8.2%; Score 167.5; DB 2; Length 747;  
 Best Local Similarity 22.5%; Pred. No. 0.0097;  
 Matches 100; Conservative 62; Mismatches 146; Indels 137; Gaps 24;

QY 7 FQWLSPLSLVAATQOLVAQPNESLPTVELPVI-----TIDSGMALANR----- 53  
 DB 42 FVTTAIVLIGIAASPAASQATADASATA-LEPIVIGGAASDSKADRTSVAANKSSAATK 100  
 QY 54 ----ITOMPHTTKYIEEQIOEQATGSRQLADVMAQLISLVSSGTTSGFQTMHGRQV 109  
 DB 101 INTPLVETPRSVSVTTEKEIEOR--GAQSIIBAVRY---SAGVTTGPNG-----FDPFRD 150  
 QY 110 QPLNVPPLT---GSRDISRQ-----LNSINPNQVARIETLSGATSI-YGSGATGGLIN 159  
 DB 151 QIFIRFENITTVGDYRDSLRQPIYNGMFRTPYQIQREVIKGPVSVLYGSSPGGLVN 210  
 QY 160 IYTKSDLEEOFTRIQVHSGKLSSEGIQVQGVAGVSENGNVLARLDVD--YRTTGG 217  
 DB 211 KISKLPTEERPIHV-----GISYTKDRAQAMFDCGPISEGNDELPIYIVGL 258

```

QY 218 A-----FDANGKR--IAP-----EP-----AQTDKQSKSLSVNTNVDWQ-LDD 253
Db 259 ARHGDNFNFOADRYFLAPSFTMKDEGTSFTLYGLAQSDETDA-NVGAIITTVGKILDI 317
QY 254 KQININLALTHYNDKQDQDYADPDYGNRLAVLFGKPSLNAIKGLSLSEQPKTKSTFNINY 313
Db 318 ROS-----DPDY--DYQVKQOQIGYQFHEFDNGLTFRQ-----NLRY 354
QY 314 HHDDL-----WGNTI--NTNAYVREKGRFPYFVAPFSIAKALPILQSMNLPSATLD 363
Db 355 SHLDLRLARVLGVSSWTGTVAHRNA-----SSIRDENNVFQVDNQLAKFD 399
QY 364 A-----YTKAPOARAYGV 376
Db 400 TGPLAHTMLFGLDYNLQSNWGYI 424

RESULT 45
O68598
ID O68598 PRELIMINARY; PRT; 742 AA.
AC O68598;
DT 01-AUG-1998 (TremBLrel. 07, Created)
DT 01-AUG-1998 (TremBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE Phenolate-type ferrisiderophore receptor.
GN PIRA.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAOI;
RX MEDLINE=96210657; PubMed=8633080;
RA Ochener U.A., Vasil M.L.;
RT "Gene repression by the ferric uptake regulator in Pseudomonas
RT aeruginosa: cycle selection of iron-regulated genes.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:4409-4414(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PAOI;
RA Ochener U.A., Vasil A.I., Johnson Z., Vasil M.L.;
RT "Genetic characterization of novel siderophore receptor genes involved
RT in iron acquisition in Pseudomonas aeruginosa.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF051692; AAC06224.1; -
DR HSSP; P05825; 1FEP.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.
KW Receptor.
SQ SEQUENCE 742 AA; 80692 MW; 572C8733FF1F4C0E CRC64;

Query Match 8.1%; Score 166.5; DB 2; Length 742;
Best Local Similarity 19.9%; Pred. NO. 0.011;
Matches 93; Conservative 66; Mismatches 146; Indels 163; Gaps 19;

QY 47 GWALA--NRITOMPHTTKVIYE-----EQIQEQATG-----SRQLADYMA 84
Db 24 GQALAEDELELDEAESAESVVLGDEVVLGTAEQLKQAPGVSIITAEIDIRKRPVNDLS 83
QY 85 QLIPSL-GVSSGTTNFGQTMHGRVQ-----FLNGVPLT-----GSRD 123
Db 84 EIIRTPGVNLTCNSSGQGRGNRQIDIRGMGPENTLILVDGKPVSSRNSVRYGWRGERD 143
QY 124 ISRLNSINPNOVAIEVLVG-ATSIYSGATGGLINIVTKSDLEE----- 168
Db 144 TRGDSNWPVPEEVERIEVLRGPAARAYGSAAGVYVNIITKPTDLRGRSMTVFTNIPES 203
QY 169 -----EQFETRIGVHGSKLSSEGIGVQVGSV----- 195
Db 204 SKDGATRANFSLSGPLTEALSFRAYGSANKTDSDDTDINLIGHTVNPSRTVAGREGVNR 263

```

```

QY 196 -----AGVSENGNVLARLDVDYRTTGGAFDANGKRIAPPAQTDKQ 236
Db 264 DLSCMLSHQVTPDQVVDFEAGFSRQGIYAG---DTQNNNGT--ANTQGLADDGAETNRM 318
QY 237 DSKSLSVNTNVDWQDLDKQININLALTHYNDKQDQDYADPDYGNRLAVLFGKPSLNAIKGL 296
Db 319 YRENYAITHNGTWSFGTSRFV---AQYD---STRNOPSGGGLAGSVEGQ---IGADRSF 368
QY 297 SLSE-OPKTKSTFNINVHHDDLWGNTINTNAYVREKGRFPYFVAPFSIAKALPILQSM 355
Db 369 SASKLENRYLSELNLP--ALFEQVLTGAEWNKE-----TL 405
QY 356 NLPSATLDAYT---KAPQARAYGVLSKESKAELVGRVFNLPKPKALF 400
Db 406 NDPSSLKQGFVGSDSLPGTTPAAGSRSPKSKAEI-----RALY 442

```

Search completed: December 25, 2002, 20:20:12  
Job time : 92 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 25, 2002, 20:22:08 ; Search time 30 Seconds  
(without alignments)  
1776.675 Million cell updates/sec

Title: US-09-889-746-2

Perfect score: 400  
Sequence: 1 MHSHPQWLSPLISAVT.....SKAEVIGRVNINPKRALF 400

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

A.GeneSeq.101002.\*  
1: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.\*  
2: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.\*  
3: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.\*  
4: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.\*  
5: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.\*  
6: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.\*  
7: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.\*  
8: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.\*  
9: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.\*  
10: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.\*  
11: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.\*  
12: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.\*  
13: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.\*  
14: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.\*  
15: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.\*  
16: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.\*  
17: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.\*  
18: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.\*  
19: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.\*  
20: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.\*  
21: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.\*  
22: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.\*  
23: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	400	100.0	400	21	AA03761
2	9	2.2	813	22	AA03758
3	8	2.0	325	19	AAW76196
4	8	2.0	325	23	AAW49565
5	8	2.0	415	22	ABG18070
6	8	2.0	1037	22	ABG04999
7	7	1.8	52	22	AAU63145
8	7	1.8	64	23	ABP33812
9	7	1.8	71	22	AAW78030
10	7	1.8	87	23	ABW53355

11	7	1.8	88	22	AA046371	H. pylori HPC188 p
12	7	1.8	89	22	ABG18574	Novel human diago
13	7	1.8	89	22	ABG18598	Novel human diago
14	7	1.8	96	21	AA032606	Eucalyptus grandis
15	7	1.8	96	21	AA033167	Eucalyptus grandis
16	7	1.8	96	21	AA030628	Human secreted pro
17	7	1.8	113	20	AAV36092	Extended human sec
18	7	1.8	114	23	AAW54592	Superoxide dismuta
19	7	1.8	126	20	AAV11637	Human 5' ESR seque
20	7	1.8	145	22	ABG20115	Novel human diago
21	7	1.8	156	18	AAW20959	H. pylori cytoplasm
22	7	1.8	164	18	AAW20518	H. pylori cytoplasm
23	7	1.8	183	22	ABW58319	Drosophila melanog
24	7	1.8	187	19	AAW62792	Tryptoxanthin perox
25	7	1.8	195	21	AAW24217	Arabidopsis thalia
26	7	1.8	209	12	AAW11535	EC-SOD mutant, T20
27	7	1.8	213	12	AAW11536	EC-SOD mutant, T21
28	7	1.8	213	12	ABW57843	Drosophila melanog
29	7	1.8	215	12	AAW11537	EC-SOD mutant, T21
30	7	1.8	216	12	AAW11538	EC-SOD mutant, T21
31	7	1.8	216	12	AAW11540	EC-SOD mutant, SAT
32	7	1.8	222	12	AAW11543	EC-SOD mutant, SAT
33	7	1.8	222	12	AAW11539	EC-SOD mutant, SAT
34	7	1.8	222	12	AAW11541	EC-SOD mutant, SAT
35	7	1.8	222	12	AAW11542	EC-SOD mutant, GI.
36	7	1.8	222	16	AAW64910	Human extracellular
37	7	1.8	223	13	AAW25161	EC-SOD. Synthetic
38	7	1.8	226	21	AAW24251	Arabidopsis thalia
39	7	1.8	235	23	ABP39982	Staphylococcus epi
40	7	1.8	240	8	AAW70702	Sequence of extrac
41	7	1.8	240	15	AAW61338	Human extracellular
42	7	1.8	240	16	AAW72413	Human EC-SOD. Hom
43	7	1.8	240	18	AAW32599	Human EC-SOD prote
44	7	1.8	240	20	AAW82447	Human EC-SOD protei
45	7	1.8	240	21	AAW94782	Human superoxide d
46	7	1.8	240	22	AAW62417	Human EC-SOD prote
47	7	1.8	240	22	AAW81063	Human extracellular
48	7	1.8	240	23	AAW52474	Superoxide dismuta
49	7	1.8	276	18	AAW20242	H. pylori transpor
50	7	1.8	276	18	AAW24615	H. pylori transpor
51	7	1.8	276	18	AAW48710	Propionibacterium
52	7	1.8	285	22	AAW33940	Staphylococcus aur
53	7	1.8	286	21	AAW24216	Arabidopsis thalia
54	7	1.8	293	22	AAW36552	Staphylococcus aur
55	7	1.8	293	22	AAW37206	Staphylococcus aur
56	7	1.8	293	22	AAW82807	S. epidermidis ope
57	7	1.8	298	23	ABP38472	Staphylococcus epi
58	7	1.8	300	22	ABW11897	Novel human diago
59	7	1.8	302	22	AAW87798	Mouse T2R04 amino
60	7	1.8	303	15	AAW60054	Diofilaria immitis
61	7	1.8	311	22	AAW03721	Murine chordin-lik
62	7	1.8	314	22	AAW71970	Human olfactory re
63	7	1.8	314	22	AAW72573	Human OR-like poly
64	7	1.8	322	22	AAW65019	Human secreted pro
65	7	1.8	328	22	AAW03733	Murine mature chor
66	7	1.8	328	14	AAW39298	Maize dwarf mosaic
67	7	1.8	333	22	AAW03720	Murine chordin-lik
68	7	1.8	333	15	AAW53752	Seven transmembran
69	7	1.8	339	18	AAW07617	Human G-protein th
70	7	1.8	339	19	AAW48733	Human R12 seven tr
71	7	1.8	339	21	AAW21697	Human 7TM receptor
72	7	1.8	339	21	AAV90618	Human G protein-co
73	7	1.8	339	21	AAV90652	Human mutant G pro
74	7	1.8	339	22	AAW82853	Human P2Y-like GPC
75	7	1.8	339	23	AAW98065	Human cytochrome l
76	7	1.8	339	23	AAW91239	Human 7 transmembr
77	7	1.8	345	22	AAW03732	Murine chordin-lik
78	7	1.8	360	22	AAW03723	Rat chordin-lik (
79	7	1.8	382	22	AAW03722	Rat chordin-lik (
80	7	1.8	382	19	AAW98237	H. pylori GPO 699
81	7	1.8	386	19	AAW71489	Helicobacter poly
82	7	1.8	394	18	AAW20797	H. pylori transpor
83	7	1.8	412	23	ABW97567	Novel human protei

84	7	1.8	416	22	AAE07964	Arabidopsis thalia	157	6	1.5	25	16	AAE64679	HPF3 peptide deriv
85	7	1.8	416	22	AAU04474	Abcisic acid resp	158	6	1.5	25	20	AAE33515	Human p75NTR depen
86	7	1.8	416	22	AAE04615	A. thaliana abscis	159	6	1.5	25	22	AAE55280	Anti-HPV3 F1 regio
87	7	1.8	416	22	AAU01869	Abcisic acid resp	160	6	1.5	26	16	AAE64680	HPF3 peptide deriv
88	7	1.8	416	22	AAU00268	Abcisic acid resp	161	6	1.5	26	21	AAE11962	Human cathepsin Y
89	7	1.8	426	22	AAE03729	Human CHLd5 mature	162	6	1.5	26	22	AAE55281	Anti-HPV3 F1 regio
90	7	1.8	429	22	AAE03728	Human CHLd5 mature	163	6	1.5	27	16	AAE64681	HPF3 peptide deriv
91	7	1.8	431	22	AAE03726	Human chordin-like	164	6	1.5	27	16	AAE64648	HPF3 peptide deriv
92	7	1.8	434	22	AAE03725	Human chordin-like	165	6	1.5	27	22	AAE55250	Anti-HPV3 F1 regio
93	7	1.8	442	20	AAE37635	Protein involved i	166	6	1.5	27	22	AAE55282	Anti-HPV3 F1 regio
94	7	1.8	447	22	AAE03727	Human CHLd5 precu	167	6	1.5	28	16	AAE64682	HPF3 peptide deriv
95	7	1.8	452	22	AAE12890	Human chordin-like	168	6	1.5	28	16	AAE64649	HPF3 peptide deriv
96	7	1.8	452	22	AAE03724	Human chordin-like	169	6	1.5	28	20	AAE33508	Human p75NTR deriv
97	7	1.8	453	22	AAE64994	Human secreted pro	170	6	1.5	28	20	AAE33509	Human p75NTR deriv
98	7	1.8	454	22	ABG30043	Novel human diagno	171	6	1.5	28	20	AAE33510	Human p75NTR deriv
99	7	1.8	457	21	AAE92777	Human chordin rela	172	6	1.5	28	20	AAE33513	Human p75NTR depen
100	7	1.8	457	21	AAE53035	Human secreted pro	173	6	1.5	28	20	AAE33524	Human p75NTR depen
101	7	1.8	489	21	AAE56573	Human prostate can	174	6	1.5	28	20	AAE33525	Human p75NTR depen
102	7	1.8	489	22	AAE65027	Gene #1 associated	175	6	1.5	28	20	AAE33525	Human secreted pro
103	7	1.8	587	19	AAE11082	H. pylori ORF hp4p	176	6	1.5	28	22	AAE55251	Anti-HPV3 F1 regio
104	7	1.8	587	22	AAE46323	H. pylori HPS188 p	177	6	1.5	28	22	AAE55283	Anti-HPV3 F1 regio
105	7	1.8	648	22	AAU34496	E. coli cellular p	178	6	1.5	29	16	AAE64683	HPF3 peptide deriv
106	7	1.8	648	22	AAE98282	Escherichia coli p	179	6	1.5	29	16	AAE64650	HPF3 peptide deriv
107	7	1.8	674	23	ABE33334	Herbicidally activ	180	6	1.5	29	21	AAE25729	Human secreted pro
108	7	1.8	734	21	AAE32174	Arabidopsis thalia	181	6	1.5	29	22	AAE55252	Anti-HPV3 F1 regio
109	7	1.8	751	21	AAE39904	Arabidopsis thalia	182	6	1.5	29	22	AAE55284	Anti-HPV3 F1 regio
110	7	1.8	756	23	ABE33267	Herbicidally activ	183	6	1.5	30	16	AAE64684	HPF3 peptide deriv
111	7	1.8	764	21	AAE32173	Arabidopsis thalia	184	6	1.5	30	16	AAE64651	HPF3 peptide deriv
112	7	1.8	781	21	AAE39903	Arabidopsis thalia	185	6	1.5	30	20	AAE10825	Amino acid sequenc
113	7	1.8	820	21	AAE322172	Arabidopsis thalia	186	6	1.5	30	22	ABG09429	Novel human diagno
114	7	1.8	837	21	AAE39902	Arabidopsis thalia	187	6	1.5	30	22	AAE55253	Anti-HPV3 F1 regio
115	7	1.8	862	22	AAU41876	Propionibacterium	188	6	1.5	30	22	AAE55285	Anti-HPV3 F1 regio
116	7	1.8	994	22	ABG19837	Novel human diagno	189	6	1.5	31	16	AAE64652	HPF3 peptide deriv
117	7	1.8	1169	22	ABG14729	Novel human diagno	190	6	1.5	31	16	AAE64685	HPF3 peptide deriv
118	7	1.8	1397	22	AAE39202	Human polypeptide	191	6	1.5	31	22	AAE55254	Anti-HPV3 F1 regio
119	7	1.8	1416	22	ABE66066	Drosophila melanog	192	6	1.5	31	22	AAE55286	Anti-HPV3 F1 regio
120	7	1.8	1416	22	AAU40988	Human polypeptide	193	6	1.5	32	16	AAE64653	HPF3 peptide deriv
121	7	1.8	2439	22	ABE68896	Drosophila melanog	194	6	1.5	32	16	AAE64686	HPF3 peptide deriv
122	7	1.8	2763	18	AAU10344	Maize dwarf mosaic	195	6	1.5	32	20	AAE11629	Human 5' EST secre
123	6	1.5	10	23	ABP47106	Human BlyS binding	196	6	1.5	32	22	AAE55255	Anti-HPV3 F1 regio
124	6	1.5	11	20	AAE33514	Human p75NTR depen	197	6	1.5	32	22	AAE55287	Anti-HPV3 F1 regio
125	6	1.5	14	16	AAE64668	HPF3 peptide deriv	198	6	1.5	33	11	AAE05848	Reagent of GTP-bin
126	6	1.5	14	20	AAE33484	Human p75NTR deriv	199	6	1.5	33	11	AAE09313	Reagent of GTP-bin
127	6	1.5	14	20	AAE33485	Human p75NTR deriv	200	6	1.5	33	16	AAE64654	HPF3 peptide deriv
128	6	1.5	14	20	AAE33486	Human p75NTR deriv	201	6	1.5	33	16	AAE64687	HPF3 peptide deriv
129	6	1.5	14	20	AAE33512	Human p75NTR depen	202	6	1.5	33	22	AAE55256	Anti-HPV3 F1 regio
130	6	1.5	14	22	AAE55269	Anti-HPV3 F1 regio	203	6	1.5	33	22	AAE55288	Anti-HPV3 F1 regio
131	6	1.5	15	16	AAE64669	HPF3 peptide deriv	204	6	1.5	34	16	AAE64655	HPF3 peptide deriv
132	6	1.5	15	22	AAE55270	Anti-HPV3 F1 regio	205	6	1.5	34	16	AAE64688	HPF3 peptide deriv
133	6	1.5	16	16	AAE64670	HPF3 peptide deriv	206	6	1.5	34	21	AAE89021	Core polypeptide f
134	6	1.5	16	22	AAE55271	Anti-HPV3 F1 regio	207	6	1.5	34	22	ABE00380	Viral Dp178/107-11
135	6	1.5	17	16	AAE64671	HPF3 peptide deriv	208	6	1.5	34	22	ABE01854	Viral core polypep
136	6	1.5	17	22	AAE55272	Anti-HPV3 F1 regio	209	6	1.5	34	22	AAU12929	Dp178-like/Dp107-1
137	6	1.5	18	21	AAE23520	Human negative elo	210	6	1.5	34	22	AAE77376	Core polypeptide f
138	6	1.5	18	22	AAE55273	Anti-HPV3 F1 regio	211	6	1.5	34	22	AAE55257	Anti-HPV3 F1 regio
139	6	1.5	19	16	AAE64673	HPF3 peptide deriv	212	6	1.5	34	22	AAE55289	Anti-HPV3 F1 regio
140	6	1.5	19	22	AAE55274	Anti-HPV3 F1 regio	213	6	1.5	35	16	AAE64656	HPF3 peptide, from
141	6	1.5	20	14	AAE39833	E1 peptide RV-EP9	214	6	1.5	35	19	AAE38328	Paramyxovirus fusi
142	6	1.5	20	16	AAE64674	HPF3 peptide deriv	215	6	1.5	35	21	AAE88834	Core polypeptide f
143	6	1.5	20	22	AAE55275	Anti-HPV3 F1 regio	216	6	1.5	35	21	AAE88835	Core polypeptide f
144	6	1.5	21	16	AAE64675	HPF3 peptide deriv	217	6	1.5	35	21	AAE88836	Core polypeptide f
145	6	1.5	21	22	AAE55276	Anti-HPV3 F1 regio	218	6	1.5	35	21	AAE88837	Core polypeptide f
146	6	1.5	22	16	AAE64676	HPF3 peptide deriv	219	6	1.5	35	21	AAE88838	Core polypeptide f
147	6	1.5	22	20	AAE11782	Human 5' EST secre	220	6	1.5	35	21	AAE88839	Core polypeptide f
148	6	1.5	22	21	AAE29833	Human secreted pro	221	6	1.5	35	21	AAE88840	Core polypeptide f
149	6	1.5	22	22	AAE55277	Anti-HPV3 F1 regio	222	6	1.5	35	21	AAE88841	Core polypeptide f
150	6	1.5	23	16	AAE64677	HPF3 peptide deriv	223	6	1.5	35	21	AAE88842	Core polypeptide f
151	6	1.5	23	18	AAE22200	Endogenous TCR alp	224	6	1.5	35	21	AAE88843	Core polypeptide f
152	6	1.5	23	22	AAE55278	Anti-HPV3 F1 regio	225	6	1.5	35	21	AAE88844	Core polypeptide f
153	6	1.5	24	16	AAE64678	HPF3 peptide deriv	226	6	1.5	35	21	AAE88845	Core polypeptide f
154	6	1.5	24	19	AAE38327	Paramyxovirus fusi	227	6	1.5	35	21	AAE88846	Core polypeptide f
155	6	1.5	24	22	AAE55279	Anti-HPV3 F1 regio	228	6	1.5	35	21	AAE88847	Core polypeptide f
156	6	1.5	24	23	AAE19019	Human B7-H3.1 pep	229	6	1.5	35	21	AAE88848	Core polypeptide f



376	6	1.5	35	22	AAU14088	DP107-like peptide	449	6	1.5	35	22	AAB54830	HPV antiviral act
377	6	1.5	35	22	AAU14089	DP107-like peptide	450	6	1.5	35	22	AAB54831	HPV antiviral act
378	6	1.5	35	22	AAU14090	DP107-like peptide	451	6	1.5	35	22	AAB55258	Anti-HPV3 P1 regio
379	6	1.5	35	22	AAU14091	DP107-like peptide	452	6	1.5	35	22	AAB55381	Anti-HPV3 peptide
380	6	1.5	35	22	AAU14092	DP107-like peptide	453	6	1.5	35	22	AAB55382	Anti-HPV3 peptide
381	6	1.5	35	22	AAB92275	Virus related pept	454	6	1.5	35	22	AAB55383	Anti-HPV3 peptide
382	6	1.5	35	22	AAB92276	Virus related pept	455	6	1.5	35	22	AAB55384	Anti-HPV3 peptide
383	6	1.5	35	22	AAB92277	Virus related pept	456	6	1.5	35	22	AAB55385	Anti-HPV3 peptide
384	6	1.5	35	22	AAB92278	Virus related pept	457	6	1.5	35	22	AAB55386	Anti-HPV3 peptide
385	6	1.5	35	22	AAB92279	Virus related pept	458	6	1.5	35	22	AAB55387	Anti-HPV3 peptide
386	6	1.5	35	22	AAB92280	Virus related pept	459	6	1.5	35	22	AAB55388	Anti-HPV3 peptide
387	6	1.5	35	22	AAB92281	Virus related pept	460	6	1.5	35	22	AAB55389	Anti-HPV3 peptide
388	6	1.5	35	22	AAB92282	Virus related pept	461	6	1.5	35	22	AAB55390	Anti-HPV3 peptide
389	6	1.5	35	22	AAB92283	Virus related pept	462	6	1.5	35	22	AAB55391	Anti-HPV3 peptide
390	6	1.5	35	22	AAB92284	Virus related pept	463	6	1.5	35	22	AAB55392	Anti-HPV3 peptide
391	6	1.5	35	22	AAB92285	Virus related pept	464	6	1.5	35	22	AAB55393	Anti-HPV3 peptide
392	6	1.5	35	22	AAB92286	Virus related pept	465	6	1.5	35	22	AAB55394	Anti-HPV3 peptide
393	6	1.5	35	22	AAB92287	Virus related pept	466	6	1.5	35	22	AAB55395	Anti-HPV3 peptide
394	6	1.5	35	22	AAB92288	Virus related pept	467	6	1.5	35	22	AAB55396	Anti-HPV3 peptide
395	6	1.5	35	22	AAB92289	Virus related pept	468	6	1.5	36	21	AAH89477	Core polypeptide f
396	6	1.5	35	22	AAB92290	Virus related pept	469	6	1.5	36	21	AAH89478	Core polypeptide f
397	6	1.5	35	22	AAB77189	Core polypeptide T	470	6	1.5	36	22	ABG16710	Novel human diagno
398	6	1.5	35	22	AAB77190	Core polypeptide T	471	6	1.5	36	22	AAU13431	DP178-like/DP107-1
399	6	1.5	35	22	AAB77191	Core polypeptide T	472	6	1.5	36	22	AAU13432	DP178-like/DP107-1
400	6	1.5	35	22	AAB77192	Core polypeptide T	473	6	1.5	37	21	AAU56344	Arabidopsis thalia
401	6	1.5	35	22	AAB77193	Core polypeptide T	474	6	1.5	37	22	AAW74159	Human bone marrow
402	6	1.5	35	22	AAB77194	Core polypeptide T	475	6	1.5	37	22	ABG44092	Human peptide enco
403	6	1.5	35	22	AAB77195	Core polypeptide T	476	6	1.5	40	22	ABG40733	Peptide #8239 enco
404	6	1.5	35	22	AAB77196	Core polypeptide T	477	6	1.5	40	22	AAH34499	Peptide #8536 enco
405	6	1.5	35	22	AAB77197	Core polypeptide T	478	6	1.5	41	19	AAV20204	Human beta-amyloid
406	6	1.5	35	22	AAB77198	Core polypeptide T	479	6	1.5	41	22	ABH10587	Human pancreatic c
407	6	1.5	35	22	AAB77199	Core polypeptide T	480	6	1.5	41	22	AAU20186	Human novel endocr
408	6	1.5	35	22	AAB77200	Core polypeptide T	481	6	1.5	42	22	ABG27934	Human peptide #585
409	6	1.5	35	22	AAB77201	Core polypeptide T	482	6	1.5	42	22	ABH33108	Peptide #614 enco
410	6	1.5	35	22	AAB77202	Core polypeptide T	483	6	1.5	42	22	ABG40970	Peptide #8476 enco
411	6	1.5	35	22	AAB77203	Core polypeptide T	484	6	1.5	42	22	ABH18574	Protein #573 enco
412	6	1.5	35	22	AAB77204	Core polypeptide T	485	6	1.5	42	22	ABH25077	Protein #7076 enco
413	6	1.5	35	22	AAB77205	Core polypeptide T	486	6	1.5	42	22	AAH53905	Human brain expres
414	6	1.5	35	22	AAB77206	Core polypeptide T	487	6	1.5	42	22	AAH61830	Human brain expres
415	6	1.5	35	22	AAB77207	Core polypeptide T	488	6	1.5	42	22	AAH66293	Human bone marrow
416	6	1.5	35	22	AAB77208	Core polypeptide T	489	6	1.5	42	22	AAH74632	Human bone marrow
417	6	1.5	35	22	AAB77209	Core polypeptide T	490	6	1.5	42	22	AAH14161	Peptide #595 enco
418	6	1.5	35	22	AAB77210	Core polypeptide T	491	6	1.5	42	22	AAH20366	Peptide #680 enco
419	6	1.5	35	22	AAB77541	Core polypeptide T	492	6	1.5	42	22	AAH26569	Peptide #606 enco
420	6	1.5	35	22	AAB77854	Core polypeptide T	493	6	1.5	42	22	AAH34745	Peptide #8782 enco
421	6	1.5	35	22	AAB77855	Core polypeptide T	494	6	1.5	42	22	AAH01900	Peptide #582 enco
422	6	1.5	35	22	AAB77867	Core polypeptide T	495	6	1.5	42	23	ABG35940	Human peptide enco
423	6	1.5	35	22	AAB77871	Core polypeptide T	496	6	1.5	42	23	ABG44458	Human peptide enco
424	6	1.5	35	22	AAB77872	Core polypeptide T	497	6	1.5	43	23	ABG44458	Amyloid beta forma
425	6	1.5	35	22	AAB77873	Core polypeptide T	498	6	1.5	45	14	AAH31370	TAP 29 protein. T
426	6	1.5	35	22	AAB77874	Core polypeptide T	499	6	1.5	45	23	ABG68725	Human prostate spe
427	6	1.5	35	22	AAB77875	Core polypeptide T	500	6	1.5	46	22	ABP35816	Plk2 positions pre
428	6	1.5	35	22	AAB77876	Core polypeptide T	501	6	1.5	46	22	ABP29134	Streptococcus poly
429	6	1.5	35	22	AAB77877	Core polypeptide T	502	6	1.5	47	22	ABH44225	Peptide #11731 enc
430	6	1.5	35	22	AAB77878	Core polypeptide T	503	6	1.5	47	22	ABH27101	Protein #9100 enco
431	6	1.5	35	22	AAB77879	Core polypeptide T	504	6	1.5	47	22	AAH65257	Human brain expres
432	6	1.5	35	22	AAB77886	Core polypeptide T	505	6	1.5	47	22	AAH77962	Human bone marrow
433	6	1.5	35	22	AAB77897	Core polypeptide T	506	6	1.5	47	22	AAH21856	Peptide #8290 enco
434	6	1.5	35	22	AAB77915	Core polypeptide T	507	6	1.5	47	22	AAH38180	Peptide #12217 enc
435	6	1.5	35	22	AAB54816	HPV antiviral act	508	6	1.5	47	23	ABG46875	Human peptide enco
436	6	1.5	35	22	AAB54817	HPV antiviral act	509	6	1.5	48	19	AAH21219	Human bcl2 proto-o
437	6	1.5	35	22	AAB54818	HPV antiviral act	510	6	1.5	50	22	ABH40235	Peptide #7741 enco
438	6	1.5	35	22	AAB54819	HPV antiviral act	511	6	1.5	50	22	ABH42086	Peptide #9592 enco
439	6	1.5	35	22	AAB54820	HPV antiviral act	512	6	1.5	50	22	ABH24665	Protein #6664 enco
440	6	1.5	35	22	AAB54821	HPV antiviral act	513	6	1.5	50	22	AAH61021	Human brain expres
441	6	1.5	35	22	AAB54822	HPV antiviral act	514	6	1.5	50	22	AAH62966	Human brain expres
442	6	1.5	35	22	AAB54823	HPV antiviral act	515	6	1.5	50	22	AAH73721	Human bone marrow
443	6	1.5	35	22	AAB54824	HPV antiviral act	516	6	1.5	50	22	AAH75778	Human bone marrow
444	6	1.5	35	22	AAB54825	HPV antiviral act	517	6	1.5	50	22	AAH88119	Human immune/haema
445	6	1.5	35	22	AAB54826	HPV antiviral act	518	6	1.5	50	22	AAH33908	Peptide #7945 enco
446	6	1.5	35	22	AAB54827	HPV antiviral act	519	6	1.5	50	22	AAH35888	Peptide #9925 enco
447	6	1.5	35	22	AAB54828	HPV antiviral act	520	6	1.5	50	23	ABG43599	Human peptide enco
448	6	1.5	35	22	AAB54829	HPV antiviral act	521	6	1.5	50	23	ABG45352	Human peptide enco



522	6	1.5	51	22	AAU46480	Protonibacterium	595	6	1.5	79	21	AAU00578	Human secreted pro
523	6	1.5	51	23	ABP29515	Streptococcus poly	596	6	1.5	79	23	ABP31439	Human ORF412 prote
524	6	1.5	52	22	AAU47848	Protonibacterium	597	6	1.5	79	23	ABP08686	Human ORFX protein
525	6	1.5	52	22	AAU63143	Protonibacterium	598	6	1.5	80	21	AAU64652	Human 5' EST relat
526	6	1.5	52	23	ABP03557	Human ORFX protein	599	6	1.5	80	22	ABP15036	Human nervous syst
527	6	1.5	53	20	AAU11545	Human 5' EST seque	600	6	1.5	81	22	AAU90814	C glutamicum prote
528	6	1.5	53	22	AAU62865	Protonibacterium	601	6	1.5	81	22	AAU91467	C glutamicum prote
529	6	1.5	53	23	ABP49606	Listeria monocytog	602	6	1.5	82	22	AAU011491	Human polypeptide
530	6	1.5	54	21	ABP33383	Human secreted pro	603	6	1.5	82	22	AAU012587	Human polypeptide
531	6	1.5	54	23	ABP07649	Human ORFX protein	604	6	1.5	83	22	AAU75390	Human colon cancer
532	6	1.5	55	20	AAU11995	Human 5' EST seque	605	6	1.5	84	18	AAU13263	Lactococcus amyli
533	6	1.5	55	22	ABG00578	Novel human diagno	606	6	1.5	84	22	AAU47895	Protonibacterium
534	6	1.5	55	22	ABG00581	Novel human diagno	607	6	1.5	84	22	ABG03656	Human musculoskele
535	6	1.5	56	17	AAW00209	Human nerve growth	608	6	1.5	84	22	AAU74615	H. pylori 5-methyl
536	6	1.5	56	22	ABH28984	Peptide #1635 enco	609	6	1.5	84	23	ABP05957	Human ORFX protein
537	6	1.5	56	22	ABH34152	Peptide #1658 enco	610	6	1.5	85	22	AAU55223	Protonibacterium
538	6	1.5	56	22	ABH19596	Protein #1595 enco	611	6	1.5	85	23	ABH81753	Tumour necrosis fa
539	6	1.5	56	22	AAU14070	DP107-like peptide	612	6	1.5	85	23	ABP42710	Human ovariin anti
540	6	1.5	56	22	AAU63931	Amino acid sequenc	613	6	1.5	86	22	AAU50035	Protonibacterium
541	6	1.5	56	22	AAU54944	Human brain expres	614	6	1.5	86	22	ABH28185	Human peptide #836
542	6	1.5	56	22	AAU67324	Human bone marrow	615	6	1.5	86	22	ABH33360	Peptide #866 enco
543	6	1.5	56	22	AAU15159	Peptide #1593 enco	616	6	1.5	86	22	ABH18820	Protein #819 enco
544	6	1.5	56	22	AAU27616	Peptide #1653 enco	617	6	1.5	86	22	AAU31143	Novel human secret
545	6	1.5	56	22	AAU02897	Peptide #1579 enco	618	6	1.5	86	22	AAU54146	Human brain expres
546	6	1.5	56	22	AAU92274	Virus related pept	619	6	1.5	86	22	AAU66539	Human bone marrow
547	6	1.5	56	22	AAU54815	HPV antiviral act	620	6	1.5	86	22	AAU14412	Peptide #846 enco
548	6	1.5	56	23	ABG36967	Human peptide enco	621	6	1.5	86	22	AAU26825	Peptide #821 enco
549	6	1.5	57	22	AAU59666	Protonibacterium	622	6	1.5	86	22	AAU02139	Peptide #862 enco
550	6	1.5	57	22	AAU05925	Human polypeptide	623	6	1.5	86	22	ABG36191	Novel human secret
551	6	1.5	58	22	AAU08460	Novel human connec	624	6	1.5	86	23	ABP38971	Human peptidase enco
552	6	1.5	59	20	AAU36318	Human secreted pro	625	6	1.5	86	23	ABP34739	Staphylococcus epi
553	6	1.5	60	23	ABP09887	Human ORFX protein	626	6	1.5	87	22	AAU59230	Human EPH family T
554	6	1.5	61	22	AAU62812	Protonibacterium	627	6	1.5	87	22	AAU82139	Protonibacterium
555	6	1.5	61	22	AAU02022	Human polypeptide	628	6	1.5	87	22	AAU09011	Human p75 neurotro
556	6	1.5	61	23	ABP02894	Human ORFX protein	629	6	1.5	88	22	AAU11663	Human ORFX protein
557	6	1.5	62	22	AAU53331	Protonibacterium	630	6	1.5	90	21	AAU41682	Human polypeptide
558	6	1.5	62	22	AAU85897	Human immune/haema	631	6	1.5	91	22	ABG14107	Human ORFX ORF146
559	6	1.5	62	22	AAU88956	Human immune/haema	632	6	1.5	91	22	ABH16884	Novel human diagno
560	6	1.5	63	21	AAU00413	Human secreted pro	633	6	1.5	91	23	ABP03305	Human nervous syst
561	6	1.5	63	21	ABP01332	Human ORFX protein	634	6	1.5	92	20	AAU97871	Human ORFX protein
562	6	1.5	64	21	AAU55703	Arbidopsis thalia	635	6	1.5	92	22	AAU46841	Human ORFX protein
563	6	1.5	64	22	ABH68187	Drosophila melanog	636	6	1.5	92	22	AAU53320	Protonibacterium
564	6	1.5	64	22	AAU57364	Protonibacterium	637	6	1.5	92	22	AAU29246	Human PRO polypept
565	6	1.5	65	21	AAU03500	Human secreted pro	638	6	1.5	92	22	AAU84535	Human immune/haema
566	6	1.5	65	22	ABG09975	Novel human diagno	639	6	1.5	92	22	AAU50921	Human PRO1157 prot
567	6	1.5	66	23	ABP10772	Human ORFX protein	640	6	1.5	92	22	AAU50960	Human secreted pro
568	6	1.5	68	22	AAU84354	Human immune/haema	641	6	1.5	93	19	AAU75179	Human secreted pro
569	6	1.5	69	22	ABG21798	Novel human diagno	642	6	1.5	93	19	AAU75126	Secreted protein 1
570	6	1.5	69	22	AAU80672	Human haematologic	643	6	1.5	93	20	AAU59683	Hydrophobic domain
571	6	1.5	69	22	AAU96454	Human reproductive	644	6	1.5	93	21	AAU12141	Arbidopsis thalia
572	6	1.5	70	21	AAU56302	Zea mays protein f	645	6	1.5	93	21	AAU34299	Human inflammation
573	6	1.5	70	21	AAU56889	Zea mays protein f	646	6	1.5	93	21	AAU94448	Human signal pepti
574	6	1.5	70	21	AAU58956	Human reproductive	647	6	1.5	94	20	AAU12768	Human 5' EST seque
575	6	1.5	70	22	AAU95750	DNA encoding NANB	648	6	1.5	94	23	ABP51396	Human MDR1 SEQ ID
576	6	1.5	72	22	AAU12657	Human polypeptide	649	6	1.5	94	23	ABP07933	Human ORFX protein
577	6	1.5	73	22	AAU61527	Protonibacterium	650	6	1.5	95	22	ABH33410	Peptide #5061 enco
578	6	1.5	73	22	AAU65610	Protonibacterium	651	6	1.5	95	22	AAU70797	Human bone marrow
579	6	1.5	73	22	AAU89922	Human immune/haema	652	6	1.5	95	22	AAU06200	Peptide #4882 enco
580	6	1.5	73	22	AAU89922	Human colon cancer	653	6	1.5	95	22	AAU7484	Listeria monocytog
581	6	1.5	74	22	AAU43777	Protonibacterium	654	6	1.5	95	23	AAU43034	Arbidopsis thalia
582	6	1.5	74	22	AAU84155	Human immune/haema	655	6	1.5	96	21	AAU64929	Protonibacterium
583	6	1.5	74	22	ABP31554	Human ORF527 prote	656	6	1.5	96	22	ABP09043	Human ORFX protein
584	6	1.5	75	22	AAU41983	Protonibacterium	657	6	1.5	97	23	AAU55565	Spider silk protei
585	6	1.5	75	22	ABG02551	Novel human diagno	658	6	1.5	98	21	ABG21336	Novel human diagno
586	6	1.5	76	22	AAU82554	Human PACAP recept	659	6	1.5	99	22	ABP06318	Human ORFX protein
587	6	1.5	77	15	AAU82554	Nerve growth facto	660	6	1.5	100	18	AAU28269	Amino acid sequenc
588	6	1.5	77	19	AAU62177	IKK binding protei	661	6	1.5	100	21	AAU00081	Human secreted pro
589	6	1.5	79	20	AAU42562	Mouse DNA accessio	662	6	1.5	100	21	AAU65524	Human gene 72-enco
590	6	1.5	79	20	AAU97872	Arbidopsis thalia	663	6	1.5	100	22	AAU66299	Protonibacterium
591	6	1.5	79	21	AAU10078	Arbidopsis thalia	664	6	1.5	100	22	AAU86385	Human immune/haema
592	6	1.5	79	21	AAU34359	Arbidopsis thalia	665	6	1.5	100	23	ABP34288	Human ORF3261 prot
593	6	1.5	79	21	AAU34359	Arbidopsis thalia	666	6	1.5	100	23	ABP00191	Human ORFX protein
594	6	1.5	79	21	AAU34359	Arbidopsis thalia	667	6	1.5	100	23	ABP00191	Human ORFX protein

668	6	1.5	100	23	ABP10837	Human ORFX protein	741	6	1.5	134	12	AAR12236	Mouse MAb 1C11 H c
669	6	1.5	101	21	AAB53290	Human colon cancer	742	6	1.5	134	20	AAV12439	Human 5' EST seque
670	6	1.5	101	21	AAG04019	Human secreted pro	743	6	1.5	134	22	ABG27109	Novel human diagno
671	6	1.5	101	21	AAV53613	Amino acid sequenc	744	6	1.5	136	12	AAR12326	Heavy chain variab
672	6	1.5	101	22	ABG23019	Novel human diagno	745	6	1.5	138	22	ABB11196	Human PG P2a recep
673	6	1.5	102	22	ABG03474	Human musculoskele	746	6	1.5	138	22	AAO12830	Human polypeptide
674	6	1.5	102	22	ABP03717	Human protein sequ	747	6	1.5	138	23	AAU75329	Human sphingosine
675	6	1.5	102	23	ABP03717	Human ORFX protein	748	6	1.5	138	23	AAU75330	Mouse sphingosine
676	6	1.5	103	22	ABG82307	S. epidermidis ope	749	6	1.5	139	14	AAR30484	VH region of Ab to
677	6	1.5	103	22	ABG82490	S. epidermidis ope	750	6	1.5	139	22	AAO02259	Human polypeptide
678	6	1.5	103	23	ABP02855	Human ORFX protein	751	6	1.5	140	22	AAO11630	Human polypeptide
679	6	1.5	104	22	AAU53902	Propionibacterium	752	6	1.5	140	22	AAU25471	Human protein sequ
680	6	1.5	105	21	AAG35524	Arabidopsis thalia	753	6	1.5	141	22	AAW79984	Human protein sequ
681	6	1.5	107	22	AAU91786	Human immune/haema	754	6	1.5	142	22	AAW88429	Human membrane or
682	6	1.5	108	23	ABP10562	Human ORFX protein	755	6	1.5	143	20	AAU49626	Corn hexose carrie
683	6	1.5	109	20	ABP27047	Amino acid sequenc	756	6	1.5	143	21	AAG10280	Arabidopsis thalia
684	6	1.5	109	21	AB441309	Human ORFX ORF1073	757	6	1.5	143	21	AAG49081	Arabidopsis thalia
685	6	1.5	109	22	AAW86648	Human immune/haema	758	6	1.5	143	21	AAU99423	Human PRO1482 (UNQ
686	6	1.5	109	23	ABP02719	Human ORFX protein	759	6	1.5	143	22	ABG15639	Novel human diagno
687	6	1.5	110	22	ABB11625	Human membrane pro	760	6	1.5	143	22	ABG19184	Novel human diagno
688	6	1.5	111	19	AAW75889	Peptide inhibitor	761	6	1.5	143	22	ABG21621	Novel human diagno
689	6	1.5	112	22	AAW86974	Human immune/haema	762	6	1.5	143	22	AAU29193	Human PRO polypept
690	6	1.5	112	22	AAW23820	Human EST encoded	763	6	1.5	143	22	AAW66172	Protein of the inv
691	6	1.5	113	23	ABW89555	Human polypeptide	764	6	1.5	144	21	AAW53445	Human colon cancer
692	6	1.5	115	20	AAU34862	Chlamydia pneumoni	765	6	1.5	144	21	AAW28378	Arabidopsis thalia
693	6	1.5	115	20	AAU34862	Human secreted pro	766	6	1.5	144	21	AAU70420	Staphylococcus aur
694	6	1.5	115	22	AAG78237	Human hexose kinas	767	6	1.5	144	22	AAU34197	Staphylococcus aur
695	6	1.5	115	23	ABP11007	Human ORFX protein	768	6	1.5	144	22	AAU36667	Staphylococcus aur
696	6	1.5	116	21	AAU10077	Arabidopsis thalia	769	6	1.5	144	22	AAU37170	Staphylococcus aur
697	6	1.5	116	22	AAU55239	Propionibacterium	770	6	1.5	144	23	ABP30168	Streptococcus poly
698	6	1.5	117	19	AAW77226	Seca protein. Sta	771	6	1.5	145	21	ABG12011	Arabidopsis thalia
699	6	1.5	117	20	AAU32780	HIV chemokine gene	772	6	1.5	145	23	ABW47851	Listeria monocytog
700	6	1.5	117	22	AAU31673	Amino acid sequenc	773	6	1.5	146	21	AAG08377	Arabidopsis thalia
701	6	1.5	118	22	AAU64479	Propionibacterium	774	6	1.5	146	21	AAU10076	Arabidopsis thalia
702	6	1.5	118	22	AAW88850	Human immune/haema	775	6	1.5	146	21	AAG34358	Arabidopsis thalia
703	6	1.5	118	23	ABP09410	Human ORFX protein	776	6	1.5	146	21	AAW50867	Arabidopsis thalia
704	6	1.5	119	21	AAW53259	Human colon cancer	777	6	1.5	146	22	ABG23130	Novel human diagno
705	6	1.5	119	23	ABP09173	Human ORFX protein	778	6	1.5	146	22	ABG24367	Novel human diagno
706	6	1.5	120	21	AAW54053	Human pancreatic c	779	6	1.5	146	22	AAW25718	Human protein sequ
707	6	1.5	121	22	AAU87262	Novel central nerv	780	6	1.5	147	21	AAW06745	Arabidopsis thalia
708	6	1.5	121	22	AAU17273	Novel signal trans	781	6	1.5	147	21	AAW69367	Amino acid sequenc
709	6	1.5	122	22	AAU25448	Human mdct protein	782	6	1.5	147	22	ABW11725	Human secreted pro
710	6	1.5	122	22	ABW65763	Drosophila melanog	783	6	1.5	147	23	ABW94331	Chlamydia pneumoni
711	6	1.5	123	22	AAO11624	Human polypeptide	784	6	1.5	148	23	ABP02355	Human ORFX protein
712	6	1.5	124	21	AAU93639	Amino acid sequenc	785	6	1.5	149	21	AAG16041	Arabidopsis thalia
713	6	1.5	124	21	AAU87319	Human signal pepti	786	6	1.5	149	21	AAG35523	Arabidopsis thalia
714	6	1.5	124	21	AAW66707	Membrane-bound pro	787	6	1.5	150	21	AAG36431	Arabidopsis thalia
715	6	1.5	124	22	AAW20118	Human immunostimul	788	6	1.5	150	21	AAG38684	Arabidopsis thalia
716	6	1.5	124	22	AAW65230	Human PRO1068 (UNQ	789	6	1.5	150	22	AAW38684	Drosophila melanog
717	6	1.5	124	22	AAW50919	Human PRO1068 prot	790	6	1.5	150	22	AAU04068	Human IL-20 associ
718	6	1.5	124	23	ABG34042	Human Pro peptide	791	6	1.5	150	22	AAW85290	Human IL-20 recept
719	6	1.5	124	23	AAU81965	Human PRO1068. Ho	792	6	1.5	150	23	ABG67221	Human interleukin
720	6	1.5	126	20	AAU30233	Clostridium sterco	793	6	1.5	150	23	ABP38178	Staphylococcus epi
721	6	1.5	126	22	AAU22822	Human prostate can	794	6	1.5	151	20	AAU34928	Chlamydia pneumoni
722	6	1.5	126	22	AAW96128	Human reproductive	795	6	1.5	151	23	ABW07399	Human protein phos
723	6	1.5	126	23	ABP42147	Human ovarian anti	796	6	1.5	152	20	AAW60208	Human endometrium
724	6	1.5	128	21	ABG33575	Arabidopsis thalia	797	6	1.5	152	21	AAG19039	Zea mays protein f
725	6	1.5	129	22	AAW94551	Human protein sequ	798	6	1.5	152	22	ABG02070	Novel human diagno
726	6	1.5	129	23	ABP01441	Human ORFX protein	799	6	1.5	153	21	AAW57053	Human prostate can
727	6	1.5	130	21	AAW34319	Human cancer assoc	800	6	1.5	153	21	AAG16040	Arabidopsis thalia
728	6	1.5	130	22	AAU95692	Cosmid CHRIMS enco	801	6	1.5	154	20	AAU35971	Human polypeptide
729	6	1.5	130	22	ABW10230	Human cDNA SEQ ID	802	6	1.5	154	23	ABW90259	Human polypeptide
730	6	1.5	131	22	AAU18026	Human immunoglobul	803	6	1.5	155	23	ABW49472	Listeria monocytog
731	6	1.5	131	21	AAW56887	Zea mays protein f	804	6	1.5	157	19	AAW41151	RBE1 transcription
732	6	1.5	131	22	ABW63740	Drosophila melanog	805	6	1.5	157	22	ABW40107	Peptide #7613 enco
733	6	1.5	132	22	ABW32752	Peptide #258 enco	806	6	1.5	157	22	AAW60855	Human brain expres
734	6	1.5	132	22	ABW18234	Protein #233 enco	807	6	1.5	157	22	AAW73550	Human bone marrow
735	6	1.5	132	22	AAW62614	Peptide #251 enco	808	6	1.5	157	23	ABW33743	Peptide #7780 enco
736	6	1.5	132	22	AAW91117	C glutamicum prote	809	6	1.5	157	23	ABW43413	Human peptide enco
737	6	1.5	132	23	ABW35584	Human peptide enco	810	6	1.5	158	22	ABW05099	Novel human diagno
738	6	1.5	133	21	AAW43329	Human ORFX ORF1093	811	6	1.5	158	23	ABW25709	Streptococcus poly
739	6	1.5	133	21	AAW43345	Human ORFX ORF109	812	6	1.5	158	23	ABW29666	Streptococcus poly
740	6	1.5	134	12	AAR12358	Heavy chain variab	813	6	1.5	158	23	ABW30648	Streptococcus poly

814	6	1.5	159	21	AA611579	Arabidopsis thailia	887	6	1.5	191	23	AAU71849	Leishmania antigen
815	6	1.5	159	21	AA616039	Arabidopsis thailia	888	6	1.5	192	21	AA615121	Arabidopsis thailia
816	6	1.5	159	21	AA626455	Arabidopsis thailia	889	6	1.5	192	21	AA646680	Arabidopsis thailia
817	6	1.5	159	21	AA636476	Arabidopsis thailia	890	6	1.5	192	22	AA691893	C glutamicum prote
818	6	1.5	159	21	AA637500	Arabidopsis thailia	891	6	1.5	192	22	AA694776	Human protein sequ
819	6	1.5	159	23	AA648352	Leishmania monocytog	892	6	1.5	193	21	AA640597	Human ORF3
820	6	1.5	160	21	AA606744	Arabidopsis thailia	893	6	1.5	194	21	AA652344	Arabidopsis thailia
821	6	1.5	160	23	AA622053	Hepatitis C virus	894	6	1.5	194	23	AA679055	Mouse pituitary tu
822	6	1.5	160	23	AA629509	Streptococcus poly	895	6	1.5	194	23	AA619598	Murine pituitary t
823	6	1.5	161	21	AA636475	Arabidopsis thailia	896	6	1.5	194	23	AA618883	Marine pituitary t
824	6	1.5	161	23	AA689434	Human polypeptide	897	6	1.5	194	23	AA608706	Rat PTTG SEQ ID NO
825	6	1.5	161	23	AA689724	Human polypeptide	898	6	1.5	195	21	AA651755	Human secreted pro
826	6	1.5	163	21	AA626918	Small subunit of p	899	6	1.5	195	21	AA610278	Arabidopsis thailia
827	6	1.5	164	22	AA664522	Drosophila melanog	900	6	1.5	195	21	AA649079	Arabidopsis thailia
828	6	1.5	164	22	AA624159	V carteri EST enco	901	6	1.5	195	21	AA65437	Arabidopsis thailia
829	6	1.5	165	22	AA682347	S. epidermidis ope	902	6	1.5	196	21	AA615120	Arabidopsis thailia
830	6	1.5	166	16	AA671191	Human erythropoiet	903	6	1.5	196	21	AA646679	Arabidopsis thailia
831	6	1.5	166	20	AA697852	Wheat ATP-phosphor	904	6	1.5	196	22	AA693922	Arabidopsis thailia
832	6	1.5	166	21	AA605785	Arabidopsis thailia	905	6	1.5	196	22	AA635402	Human polypeptide,
833	6	1.5	166	22	AA656780	Propionibacterium	906	6	1.5	197	21	AA652343	Replication protei
834	6	1.5	166	22	AA623785	Novel human enzyme	907	6	1.5	197	22	AA691789	Arabidopsis thailia
835	6	1.5	166	22	AA621846	Novel human neopla	908	6	1.5	198	21	AA646678	C glutamicum prote
836	6	1.5	167	22	AA617913	Human nervous syst	909	6	1.5	198	22	AA659952	Arabidopsis thailia
837	6	1.5	168	15	AA653074	Mouse GRP78. Mus	910	6	1.5	198	22	AA65768	Drosophila melanog
838	6	1.5	170	22	AA615075	Novel human diagno	911	6	1.5	198	23	AA641770	Cysteine protease
839	6	1.5	170	22	AA694349	Human reproductive	912	6	1.5	199	18	AA623715	Human ovarian anti
840	6	1.5	171	21	AA646697	Arabidopsis thailia	913	6	1.5	199	19	AA64899	Dirofilaria immiti
841	6	1.5	172	21	AA605784	Arabidopsis thailia	914	6	1.5	199	21	AA655436	Rat pituitary-tumo
842	6	1.5	172	22	AA642153	Human polypeptide	915	6	1.5	199	21	AA650613	Arabidopsis thailia
843	6	1.5	173	21	AA636474	Arabidopsis thailia	916	6	1.5	199	22	AA602929	Arabidopsis thailia
844	6	1.5	174	21	AA635522	Novel human diagno	917	6	1.5	199	22	AA624049	Novel human diagno
845	6	1.5	175	22	AA611135	Novel human diagno	918	6	1.5	199	22	AA614537	Human EST encoded
846	6	1.5	175	23	AA655585	Lactococcus lactis	919	6	1.5	199	23	AA679052	Human novel protei
847	6	1.5	176	22	AA694716	Human protein sequ	920	6	1.5	199	23	AA618594	Rat pituitary tumo
848	6	1.5	176	23	AA600200	Human ORF3 protein	921	6	1.5	199	23	AA618860	Rat pituitary tumo
849	6	1.5	177	21	AA628977	Arabidopsis thailia	922	6	1.5	199	23	AA618860	Rat pituitary tumo
850	6	1.5	177	22	AA695539	Human protein sequ	923	6	1.5	200	10	AA690732	Human growth hormo
851	6	1.5	178	21	AA629185	Arabidopsis thailia	924	6	1.5	200	21	AA630443	Fish growth hormo
852	6	1.5	178	21	AA634230	Arabidopsis thailia	925	6	1.5	200	22	AA603676	Arabidopsis thailia
853	6	1.5	178	21	AA652345	Arabidopsis thailia	926	6	1.5	202	22	AA671547	Novel human diagno
854	6	1.5	179	22	AA682979	S. epidermidis ope	927	6	1.5	203	10	AA690617	Human olfactory re
855	6	1.5	180	23	AA683863	Bacillus lipase po	928	6	1.5	203	11	AA607248	Fish growth hormo
856	6	1.5	180	23	AA683864	Bacillus lipase po	929	6	1.5	203	11	AA607248	Fish growth hormo
857	6	1.5	181	21	AA610455	Arabidopsis thailia	930	6	1.5	204	10	AA695638	Drosophila melanog
858	6	1.5	181	22	AA616707	Novel human diagno	931	6	1.5	204	12	AA610912	Fish growth hormo
859	6	1.5	181	22	AA601855	Rat PDGF-associated	932	6	1.5	204	18	AA625120	Yellowtail tuna gr
860	6	1.5	182	20	AA635428	Chlamydia pneumoni	933	6	1.5	204	18	AA627340	Yellow tail/fin tu
861	6	1.5	182	21	AA611050	Arabidopsis thailia	934	6	1.5	204	20	AA649020	Korean rockfish gr
862	6	1.5	182	22	AA682130	S. epidermidis ope	935	6	1.5	204	21	AA611163	Arabidopsis thailia
863	6	1.5	183	21	AA632673	Eucalyptus grandis	936	6	1.5	205	19	AA658875	S. pneumoniae deri
864	6	1.5	183	21	AA611578	Arabidopsis thailia	937	6	1.5	205	21	AA625861	Arabidopsis thailia
865	6	1.5	183	21	AA697355	Human signal pepti	938	6	1.5	207	19	AA658051	H. contortus pHC-2
866	6	1.5	183	22	AA609626	Human gene 2 encod	939	6	1.5	207	21	AA642375	Arabidopsis thailia
867	6	1.5	183	22	AA609633	Human gene 2 encod	940	6	1.5	207	21	AA646696	Arabidopsis thailia
868	6	1.5	185	21	AA610279	Arabidopsis thailia	941	6	1.5	207	22	AA685289	Human IL-20 recept
869	6	1.5	185	21	AA634368	Arabidopsis thailia	942	6	1.5	207	22	AA652465	Human FD3 amino a
870	6	1.5	185	21	AA649080	Arabidopsis thailia	943	6	1.5	207	23	AA657220	Interleukin-20 sub
871	6	1.5	185	22	AA674588	Human colon cancer	944	6	1.5	208	22	AA633806	Scaphylococcus aur
872	6	1.5	186	20	AA697853	Wheat ATP-phosphor	945	6	1.5	211	21	AA608861	Arabidopsis thailia
873	6	1.5	186	21	AA649260	Arabidopsis thailia	946	6	1.5	211	21	AA625261	Arabidopsis thailia
874	6	1.5	186	21	AA686523	Human gene 72-enco	947	6	1.5	211	23	AA689718	Human polypeptide
875	6	1.5	186	22	AA671931	Drosophila melanog	948	6	1.5	212	18	AA618065	Rat urogenital sin
876	6	1.5	186	23	AA639896	Staphylococcus epi	949	6	1.5	212	21	AA607187	Arabidopsis thailia
877	6	1.5	187	22	AA652711	Escherichia coli p	950	6	1.5	212	21	AA609935	Arabidopsis thailia
878	6	1.5	188	20	AA629789	Thymidylate kinase	951	6	1.5	212	21	AA627646	Arabidopsis thailia
879	6	1.5	189	18	AA635904	Human haematopoiet	952	6	1.5	213	20	AA682684	Human secreted pro
880	6	1.5	189	21	AA634694	Human secreted pro	953	6	1.5	214	22	AA600580	Novel human diagno
881	6	1.5	189	22	AA611618	Human GABA transpo	954	6	1.5	214	22	AA685287	Interleukin-20 sub
882	6	1.5	189	22	AA680049	Human protein SEO	955	6	1.5	215	23	AA657218	Arabidopsis thailia
883	6	1.5	189	22	AA620096	Human hydropobic	956	6	1.5	215	21	AA619526	Arabidopsis thailia
884	6	1.5	191	19	AA670253	Leishmania antigen	957	6	1.5	215	21	AA654798	Arabidopsis thailia
885	6	1.5	191	22	AA648523	Propionibacterium	958	6	1.5	216	22	AA600950	Human B7-H3 polype
886	6	1.5	191	23	AA660905	Leishmania antigen	959	6	1.5	216	22	AA687417	Human gene 2 encod

960 6 1.5 217 22 AAU25428 Human mddt protein  
 961 6 1.5 217 22 AAB85280 IL-20RA-Ig gamma1  
 962 6 1.5 217 23 ABG67211 Interleukin-20 sub  
 963 6 1.5 218 21 AAG60612 Arabidopsis thalia  
 964 6 1.5 219 19 AAW79435 Staphylococcus aur  
 965 6 1.5 219 21 AAG42374 Arabidopsis thalia  
 966 6 1.5 219 21 AAG55435 Arabidopsis thalia  
 967 6 1.5 219 22 AAU36575 Staphylococcus aur  
 968 6 1.5 219 22 AAU37450 Staphylococcus aur  
 969 6 1.5 219 22 ABG07918 Novel human diagno  
 970 6 1.5 220 21 ABA43764 Human cancer assoc  
 971 6 1.5 220 22 AAO05752 Human polypeptide  
 972 6 1.5 221 20 AAY48547 Human breast tumou  
 973 6 1.5 221 21 AAB58269 Lung cancer associ  
 974 6 1.5 221 23 AAB85269 Human IL-20 recept  
 975 6 1.5 221 23 ABG67200 Human Interleukin-  
 976 6 1.5 221 23 AAE23354 Human IL-20RA matu  
 977 6 1.5 222 22 ABB59370 Drosophila melanog  
 978 6 1.5 222 22 AAU04853 Microspora eve  
 979 6 1.5 222 23 AAU83128 Novel secreted pro  
 980 6 1.5 223 21 AAG48694 Arabidopsis thalia  
 981 6 1.5 224 22 ABG13943 Novel human diagno  
 982 6 1.5 225 22 AAG72112 Human olfactory re  
 983 6 1.5 226 22 ABG79448 Novel human diagno  
 984 6 1.5 231 22 AAU37078 Staphylococcus aur  
 985 6 1.5 231 22 AAG92262 C glutamicum prote  
 986 6 1.5 232 21 AAG40852 Zea mays protein f  
 987 6 1.5 232 21 AAG12947 Arabidopsis thalia  
 988 6 1.5 233 21 AAG45830 Arabidopsis thalia  
 989 6 1.5 233 22 AAM86095 Human immune/haema  
 990 6 1.5 234 21 AAB51802 Gene 26 human secr  
 991 6 1.5 234 21 AAB51803 Human secreted pro  
 992 6 1.5 234 21 AAG26430 Arabidopsis thalia  
 993 6 1.5 234 21 AAG38683 Arabidopsis thalia  
 994 6 1.5 234 22 AAG10334 Novel human diagno  
 995 6 1.5 235 21 AAG52260 Arabidopsis thalia  
 996 6 1.5 235 22 AAB36214 Human immune syste  
 997 6 1.5 235 23 AAB54919 Lactococcus lactis  
 998 6 1.5 236 15 AAR47462 Truncated xylanase  
 999 6 1.5 236 16 AAR64807 PNA lectin subunit  
 1000 6 1.5 236 21 AAB28136 Senescence-like pr

ALIGNMENTS

RESULT 1  
 ID AAB07761 standard; Protein; 400 AA.

XX AAB07761;  
 AC  
 XX  
 DT 07-NOV-2000 (first entry)  
 XX  
 DE Amino acid sequence of a BASB047 polypeptide.  
 KW BASB047; BASB054; BASB068; BASB069; vaccine; bacteremia; meningitis;  
 KW Neisseria meningitidis disease; upper respiratory tract infection;  
 KW invasive bacterial disease.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO200043519-A2.  
 XX  
 PD 27-JUL-2000.  
 XX  
 PF 19-JAN-2000; 2000WO-EP00428.  
 XX  
 PR 22-JAN-1999; 99GB-0001368.  
 PR 28-JAN-1999; 99GB-0001944.  
 PR 29-JAN-1999; 99GB-0002086.  
 PR 15-FEB-1999; 99GB-0003417.  
 PR 16-FEB-1999; 99GB-0003535.

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 PA Ruelle J;  
 PI WPI; 2000-505839/45.  
 XX N-PSDB; AAA59347.  
 DR  
 DR Neisseria meningitidis BASB047, BASB054, BASB068, and BASB069 proteins,  
 PT useful for treating N. meningitidis infections, bacteremia, and  
 PT meningitis -  
 XX  
 PS Claim 3; Page 93-94; 103pp; English.  
 XX The specification describes Neisseria meningitidis polypeptides  
 CC designated BASB047, BASB054, BASB068, and BASB069. The polynucleotide  
 CC sequences can be used to create a vector to transform a host cell.  
 CC The host cell can be used to produce the polypeptide. The polynucleotides  
 CC and polypeptides can be used in vaccine compositions. The  
 CC polynucleotides, polypeptides, and antibodies directed against the  
 CC polypeptides can be used in compositions for preparation of medicaments.  
 CC The antibodies can also be used in a composition for treating humans  
 CC with Neisseria meningitidis disease. The diseases that can be treated  
 CC include upper respiratory tract infection, and invasive bacterial  
 CC diseases such as bacteremia and meningitis. The nucleic acid sequences  
 CC can be used as probes in the diagnosis of Neisseria meningitidis disease.  
 CC The present sequence represents a BASB047 polypeptide.  
 XX  
 SQ Sequence 400 AA;  
 Query Match 100.0%; Score 400; DB 2i; Length 400;  
 Best Local Similarity 100.0%; Pred. NO. 0;  
 Matches 400; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MRHSHYFOWLSPLLSVAVTQQLYAQPNESLPTVELEPVITIDKSGNALANRITOMPH 60  
 DB 1 MRHSHYFOWLSPLLSVAVTQQLYAQPNESLPTVELEPVITIDKSGNALANRITOMPH 60  
 QY 61 TKVIYEEIQEOATGSRQLADVMAQLIPSLCVSSGTTSNFQGTWHGRQVQFLNGVPLTG 120  
 DB 61 TKVIYEEIQEOATGSRQLADVMAQLIPSLCVSSGTTSNFQGTWHGRQVQFLNGVPLTG 120  
 QY 121 SRDISRQLNSINPNOVARIEVLSCGATSIYSGATGGLINIVTKSDLEEEQFETRIGVHGS 180  
 DB 121 SRDISRQLNSINPNOVARIEVLSCGATSIYSGATGGLINIVTKSDLEEEQFETRIGVHGS 180  
 QY 181 KLSSEGIGYQVGSVAGVSENGNVLARDVDYRTTGGAFDANGKRIAPEPAQTDKQDSKS 240  
 DB 181 KLSSEGIGYQVGSVAGVSENGNVLARDVDYRTTGGAFDANGKRIAPEPAQTDKQDSKS 240  
 QY 241 LSVNTNVDWQLDDKQINLALTHYNDKQDTPYAPDYGNRLAVLFGKEKPSLNAIKGLSLSE 300  
 DB 241 LSVNTNVDWQLDDKQINLALTHYNDKQDTPYAPDYGNRLAVLFGKEKPSLNAIKGLSLSE 300  
 QY 301 QPKTTKSTFNINHYHDDLGWNTINTNAYRREKGRFPFVAPFSTAKALPILQSNMLPSA 360  
 DB 301 QPKTTKSTFNINHYHDDLGWNTINTNAYRREKGRFPFVAPFSTAKALPILQSNMLPSA 360  
 QY 361 TLDAYTKAPQARAYGVLOSSESKEAEVLGRVPLNPKRALF 400  
 DB 361 TLDAYTKAPQARAYGVLOSSESKEAEVLGRVPLNPKRALF 400  
 RESULT 2  
 AAU33589  
 ID AAU33589 standard; Protein; 813 AA.  
 XX  
 AC AAU33589;  
 XX  
 DT 14-FEB-2002 (first entry)  
 XX  
 DE Pseudomonas aeruginosa cellular proliferation protein #33.  
 XX

KM Antisense; prokaryotic cellular proliferation protein;  
 KM antibiotic; antibacterial; drug design.  
 XX  
 OS Pseudomonas aeruginosa.  
 XX  
 PN W0200170955-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 21-MAR-2001; 2001WO-US09180.  
 XX  
 PR 21-MAR-2000; 2000US-191078P.  
 PR 23-MAY-2000; 2000US-206848P.  
 PR 26-MAY-2000; 2000US-207727P.  
 PR 23-OCT-2000; 2000US-242578P.  
 PR 27-NOV-2000; 2000US-253625P.  
 PR 22-DEC-2000; 2000US-257931P.  
 PR 16-FEB-2001; 2001US-269308P.  
 XX  
 PA (ELIT- ) ELITRA PHARM INC.  
 XX  
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,  
 PI Yamamoto RT, Xu HH;  
 XX  
 DR WPI; 2001-611495/70.  
 DR N-PSDB; AAS51448.  
 XX  
 PT New polynucleotides for the identification and development of  
 PT antibiotics, comprise sequences of antisense nucleic acids -  
 XX  
 PS Example 3; Seq ID No 5085; 511pp; English.  
 XX  
 CC The invention relates to antisense inhibitors of genes essential to  
 CC prokaryotic cellular proliferation, their use in identifying the  
 CC genes, their use in the discovery of novel antibiotics, the essential  
 CC genes themselves and the encoded proteins. The prokaryotes used are  
 CC *Bacteriella coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*  
 CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The  
 CC invention is also useful for the identification of potential new targets  
 CC for antibiotic development. The antisense nucleic acids can also be used  
 CC to identify proteins used in proliferation, to express these proteins,  
 CC and to obtain antibodies capable of binding to the expressed proteins.  
 CC The proteins can be used to screen compounds in rational drug discovery  
 CC programmes. The antisense nucleic acid sequence is also useful to screen  
 CC for homologous nucleic acids which are required for cell proliferation in  
 CC a wide variety of organisms. The present sequence represents an  
 CC essential prokaryotic cellular proliferation protein.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 813 AA;  
 XX  
 QY Query Match 2.2%; Score 9; DB 22; Length 813;  
 Db Best Local Similarity 100.0%; Pred. No. 6.1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 148 IYSGGATGG 156  
 Db 229 IYSGGATGG 237  
 XX  
 RESULT 3  
 ID AAM76196 standard; Protein; 325 AA.  
 AC AAM76196;  
 XX  
 XX 26-NOV-1998 (first entry)  
 DT  
 XX Actinoplanes sp. acarbose protein acbP.  
 DE  
 XX

KM Acarbose; acb cluster; acbQ; acbK; acbM; acbL; acbC; acbA; acbB; acbD;  
 KM acbE; acbG; acbF; acbH; maltodextrin transferase; dtp-glucose synthase;  
 KM dtp-glucose-4,6-dehydratase; C7 sugar cyclase; acaryosyl transferase;  
 KM alpha amylase; MaltF-like membrane protein; MaltF-like membrane protein;  
 KM MaltF-like binding protein; acarbose-7-kinase; oxidoreductase;  
 KM maltodextrin glucotransferase.  
 XX  
 OS Actinoplanes sp.  
 XX  
 PN DE19708127-A1.  
 XX  
 PD 03-SEP-1998.  
 XX  
 PF 28-FEB-1997; 97DE-1008127.  
 XX  
 PR 28-FEB-1997; 97DE-1008127.  
 XX  
 PA (FARB ) BAYER AG.  
 XX  
 PI Apeler H, Gruenger A, Distler J, Goeke K, Jarling M;  
 PI Pape H, Paz Marta DU, Piepersberg W, Schroeder W;  
 PI Struttman A;  
 XX  
 DR WPI; 1998-468360/41.  
 DR N-PSDB; AAV56642.  
 XX  
 PT New actinoplanes acarbose gene cluster - useful for increasing  
 PT acarbose production in Actinoplanes, etc  
 XX  
 PS Disclosure; Fig 3; 51pp; German.  
 XX  
 CC This sequence represents the acbP protein from the Actinoplanes acarbose  
 CC gene cluster which encodes the acbA, acbB, acbC, acbD, acbE, acbF, acbG,  
 CC acbH, acbK, acbL, acbM and acbQ proteins. The gene cluster is useful for  
 CC increasing acarbose production in Actinoplanes or other microorganisms or  
 CC in vitro synthesis systems. This protein is a putative MaltF-like  
 CC membrane protein.  
 XX  
 SQ Sequence 325 AA;  
 XX  
 QY Query Match 2.0%; Score 8; DB 19; Length 325;  
 Db Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 14 LLSVAVTQ 21  
 Db 278 LLSVAVTQ 285  
 XX  
 RESULT 4  
 ID AAM49565 standard; Protein; 325 AA.  
 AC AAM49565;  
 XX  
 XX 31-MAY-2002 (first entry)  
 DT  
 XX Actinoplanes sp acarbose synthase AcbPASP.  
 DE  
 XX Acarbose; AcbAASP; AcbBASP; AcbCAASP; AcbDAASP; AcbEASP; AcbFASP; AcbGASP;  
 KM AcbHASP; AcbIASP; AcbJASP; AcbKASP; AcbLASP; AcbMASP; AcbNASP; AcbOASP;  
 KM AcbPASP; AcbQASP; AcbRAASP; AcbSASP; AcbTASP; AcbUASP; AcbVASP; AcbWASP;  
 KM AcbXASP; AcbYASP; AcbZASP; Asp3-1; Asp3-2; Asp3-3; bioconversion;  
 KM biosynthesis; alpha-glucosidase inhibitor.  
 XX  
 OS Actinoplanes sp.  
 XX  
 PN DE10021667-A1.  
 XX  
 XX 08-NOV-2001.  
 DT  
 XX 05-MAY-2000; 2000DE-1021667.  
 PF  
 XX



XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 1037 AA;

Query Match 2.0%; Score 8; DB 22; Length 1037;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 GYVQSESK 382  
Db 105 GYVQSESK 112  
|||||

RESULT 7

ID AAU63145 standard; Protein; 52 AA.

XX AC AAU63145;

XX DT 27-FEB-2002 (first entry)

XX DE Propionibacterium acnes immunogenic protein #24041.

XX SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;  
XX uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
XX inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
XX dermatological; osteopathic; neuroprotectant.

XX OS Propionibacterium acnes.

XX PN WO200181581-A2.

XX PD 01-NOV-2001.

XX PF 20-APR-2001; 2001WO-US12865.

XX PR 21-APR-2000; 2000US-199047P.

XX PR 02-JUN-2000; 2000US-208841P.

XX PR 07-JUL-2000; 2000US-216747P.

XX PA (CORI-) CORIXA CORP.

XX PI Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

XX PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX DR WPI; 2001-616774/71.

XX DR N-PSDB; AAS59632.

XX PT Propionibacterium acnes polypeptides and nucleic acids useful for  
XX PT vaccinating against and diagnosing infections, especially useful for  
XX PT treating acne vulgaris.

PS Example 1; SEQ ID No 24340; 1069pp; English.

XX Sequences AAU9105-AAU68017 represent Propionibacterium acnes immunogenic  
XX polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA).  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 52 AA;

Query Match 1.8%; Score 7; DB 22; Length 52;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 VPLTGR 122  
Db 28 VPLTGR 34  
|||||

RESULT 8

ID ABP3812 standard; Protein; 64 AA.

XX AC ABP3812;

XX DT 08-JUL-2002 (first entry)

XX DE Human ORF2785 protein, SEQ ID NO:5570.

XX Human, ORF: open reading frame; ORF: drug screening; diagnosis;  
XX disease monitoring; cytokine; cell proliferation; cell differentiation;  
XX immune modulation; haematopoiesis regulation; tissue growth;  
XX angiogenesis; activin; inhibin; chemotactic; chemokine; haemostatic;  
XX thrombolytic; tumour inhibition; bodily characteristics; fertility;  
XX behaviour; cancer; proliferative disorder; neurological disorder;  
XX cardiovascular disease; immune system disorder; organ transplantation;  
XX tissue growth disorder; tissue regeneration disorder; diabetes mellitus;  
XX hypothyroidism; cholesterol ester storage disease; infection; vulnery;  
XX vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;  
XX neuroprotective; antithrombotic; anticoagulant; thrombolytic;  
XX cardiac; hypotensive; antihypertoid; antiinflammatory; immunomodulator;  
XX dermatological; analgesic; virucide; antibacterial; fungicide.

XX OS Homo sapiens.

XX PN WO200190366-A2.

XX PD 29-NOV-2001.

XX PF 24-MAY-2001; 2001WO-US17076.

XX PR 24-MAY-2000; 2000US-206690P.

XX PA (CURA-) CURAGEN CORP.

XX PI Leach MD, Shinkets RA;

XX DR WPI; 2002-106200/14.

DR N-PSDB; ABN77838.  
XX Novel human polypeptides and polynucleotides useful for diagnosing,  
PT preventing and treating cardiovascular disease, neurodegenerative,  
PT hyperproliferative disorders and disorders related to organ  
PT transplantation -  
XX Claim 10; Page 1643; 2508pp; English.  
XX Sequences ABP31028-ABP35561 represent 4534 novel human proteins  
CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-  
CC ABN79587 represent cDNAs encoding them. The invention also encompasses  
CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively  
CC referred to as ORFX) proteins, polynucleotides at least 85% identical to  
CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX  
CC polynucleotides, the recombinant production of ORFX proteins, antibodies  
CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and  
CC polypeptides, methods of screening for modulators of ORFX expression or  
CC activity, and methods of screening individuals for a predisposition to an  
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide  
CC range of biological activities, such as cytokine, cell proliferation,  
CC cell differentiation, immune modulation, haematopoiesis regulation,  
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/  
CC chemokinetic activity, haemostatic activity, thrombolytic activity,  
CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,  
CC and antiinfective activity, and may also be involved in the determination  
CC of bodily characteristics, fertility and behaviour. ORFX proteins,  
CC nucleic acids and antibodies may be used in the treatment of cancers,  
CC other proliferative disorders such as psoriasis and benign tumours,  
CC neurological disorders such as epilepsy and Alzheimer's disease,  
CC cardiovascular diseases, immune system disorders, disorders related to  
CC organ transplantation, disorders of tissue growth and regeneration,  
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol  
CC storage disease, and infectious diseases caused by viral, bacterial,  
CC fungal and other pathogens. ORFX nucleic acids may also be used as a  
CC source of primers and probes, in the detection of ORFX genomic sequences  
CC or transcripts, in the identification and cloning of homologous  
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX  
CC nucleic acids may additionally be used to produce transgenic animals  
CC which may be useful for studying the function and/or activity of ORFX  
CC protein, and in drug screening. The ORFX proteins may also be used as  
CC immunogens to generate specific antibodies, which are useful in the  
CC diagnosis, treatment and monitoring of ORFX-associated diseases.  
XX  
SQ Sequence 64 AA;  
Query Match 1.8%; Score 7; DB 23; Length 64;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 66 EEQIQEQ 72  
Db 31 EEQIQEQ 37  
RESULT 9  
AAG78030  
ID AAG78030 standard; Protein; 71 AA.  
AC AAG78030;  
XX 15-JAN-2002 (first entry)  
DT  
DE Piscirickettsia salmonis polypeptide clone 3/JAPA-F.  
XX Piscirickettsia salmonis; piscirickettsiosis; salmonid rickettsial;  
KW Piscirickettsia salmonis; piscirickettsiosis; salmonid rickettsial;  
XX septicemia; SRS; surface antigen; vaccine; antibacterial; fish.  
OS Piscirickettsia salmonis.  
XX  
XX Key Location/Qualifiers  
FH Misc-difference 62 /label= unknown  
FT

FT /note= "Encoded by CAN"  
XX WO200168865-A2.  
XX 20-SEP-2001.  
XX 12-MAR-2001; 2001WO-GB01055.  
XX 11-MAR-2000; 2000GB-0005838.  
PR 01-JUL-2000; 2000GB-0016080.  
PR 01-JUL-2000; 2000GB-0016082.  
PR 29-JUL-2000; 2000GB-0018599.  
XX (AQUA-) AQUA HEALTH EURO LTD.  
XX Simard N, Brouwers H, Jones S, Griffiths S, Valenzuela P;  
PI Burzio L;  
XX WPI; 2001-639050/73.  
DR N-PSDB; AAH79044.  
XX New nucleic acids encoding an amino acid sequence homologous to the  
PT surface antigen present on Piscirickettsia salmonis are useful to  
PT protect fish against piscirickettsiosis -  
XX Claim 6; Fig 10; 25pp; English.  
XX The invention relates to nucleic acid sequences and the encoded protein  
CC of a least part of the surface antigen present on Piscirickettsia  
CC salmonis for production of a vaccine with antibacterial activity to  
CC protect fish against P. salmonis which causes piscirickettsiosis, also  
CC known as salmonid rickettsial septicemia.  
XX  
SQ Sequence 71 AA;  
Query Match 1.8%; Score 7; DB 22; Length 71;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 121 SRDISRQ 127  
Db 40 SRDISRQ 46  
RESULT 10  
ABB53355  
ID ABB53355 standard; Protein; 87 AA.  
XX ABB53355;  
AC ABB53355;  
XX 16-MAY-2002 (first entry)  
DT  
DE Lactococcus lactis protein yafJ.  
XX Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.  
KW Lactococcus lactis IL1403.  
OS FR2807446-A1.  
XX PN 12-OCT-2001.  
XX 11-APR-2000; 2000FR-0004630.  
XX 11-APR-2000; 2000FR-0004630.  
XX (INRG ) INRA INST NAT RECH AGRONOMIQUE.  
XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;  
PI WPI; 2002-043418/06.  
XX New nucleotide sequence useful in the identification or Lactococcus  
PT



PT lactic acid and related species -  
 XX  
 PS Claim 6; SEQ ID No 57; 2504pp; French.  
 CC  
 CC The present invention is related to a lactic acid bacteria nucleotide  
 CC sequence (AB90521) and related proteins (AB53300-AB555621). The  
 CC nucleic acid sequence is useful in the detection and/or amplification of  
 CC lactic acid bacteria, particularly to identify lactic acid bacteria or  
 CC related species. The proteins of the invention are useful for the  
 CC bioanalysis or biodegradation of a composition of interest. The  
 CC invention helps research in lactic acid bacteria, particularly useful in the  
 CC production of yogurt and cheese.  
 CC Note: The sequence data for this patent is based on equivalent patent  
 CC WO200177334 (published 18-OCT-2001) which is available in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
 SQ Sequence 87 AA;  
 QY Query Match 1.8%; Score 7; DB 23; Length 87;  
 Best Local Similarity 100.0%; Pred. No. 74;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 91 GVSSGT 97  
 43 GVSSGT 49  
 RESULT 11  
 AAB46371  
 ID AAB46371 standard; Protein; 88 AA.  
 AC AAB46371;  
 DT 05-APR-2001 (first entry)  
 XX  
 DE H. pylori HPC188 protein.  
 XX  
 KW Microbial infection; antibacterial; Helicobacter pylori infection;  
 KW vaccine; screening.  
 XX  
 OS Helicobacter pylori.  
 XX  
 PN WO200073502-A2.  
 PD 07-DEC-2000.  
 XX  
 PF 31-MAY-2000; 2000WO-EP05024.  
 XX  
 PR 31-MAY-1999; 99DE-1024965.  
 PR 17-JUN-1999; 99DE-1027740.  
 PR 21-JUL-1999; 99DE-1034029.  
 XX  
 PA (PLAC) MAX PLANCK GES FORBILDUNG WISSENSCHAFTEN.  
 PA (CREA-) CREATOGEN GMBH.  
 PI Apfel H, Fuchs TM, Gibbs CP, Hueck CJ, Meyer TF;  
 PI MPI; 2001-049948/06.  
 DR N-PSDB; AAF25648.  
 DR  
 PT Preparing an agent for diagnosis or control of microbial infection,  
 PT useful particularly against Helicobacter, based on identification of  
 PT essential genes in defective mutants -  
 XX  
 PS Claim 37; Figure 15; 366pp; German.  
 CC This invention describes a novel preparation of an agent (A) for  
 CC detection, prevention and/or treatment of microbial infection by:  
 CC (i) identifying essential genes (I) and corresponding polypeptides  
 CC (II); (ii) identifying compounds that are directed against (II) and  
 CC inactivate the microbe; (iii) testing these for suitability for use; and  
 CC (iv) formulating selected (A). Identifying essential genes (I) comprises  
 CC preparation of gene-deficient microorganisms by conditional antisense

CC inhibition (CAI) and/or subtractive recombination mutagenesis (SRM),  
 CC then determining viability and/or survival of the deficient organisms.  
 CC The products of the invention have antibacterial activity. (A) (which may  
 CC be a nucleic acid (Ia), vector or host cell containing (Ia), derived  
 CC polypeptide (IIa), or fragments, (IIa)-specific antibodies or their  
 CC fragments or an inhibitor of (IIa)) are particularly used for diagnosis,  
 CC treatment or prevention of infection by Helicobacter pylori. Particularly  
 CC (Ia) and (IIa) are used in DNA, subunit or live vaccines. The method  
 CC identifies essential genes, including those that have homologs in other  
 CC species, so identified (A) should have a broad spectrum of activity. Many  
 CC gene-deficient cells can be screened quickly, in an automated process,  
 CC and the identified genes can be used for screening without purification.  
 SQ Sequence 88 AA;  
 QY Query Match 1.8%; Score 7; DB 22; Length 88;  
 Best Local Similarity 100.0%; Pred. No. 74;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 10 LSLPLS 16  
 21 LSLPLS 27  
 RESULT 12  
 ABG18574  
 ID ABG18574 standard; Protein; 89 AA.  
 AC ABG18574;  
 DT 18-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #18565.  
 XX  
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US08631.  
 XX  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 PI Drmanac RT, Liu C, Tang YT;  
 PI WPI; 2001-639362/73.  
 DR N-PSDB; AAS82761.  
 DR  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 20; SEQ ID No 48933; 103pp; English.  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 89 AA;  
Query Match 1.8%; Score 7; DB 22; Length 89;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 88 PSLGVSS 94  
|||  
Db 44 PSLGVSS 50

RESULT 13  
ABG18598  
ID ABG18598 standard; Protein; 89 AA.

XX AC ABG18598;

DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #18589.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS82785.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity.

XX PS Claim 20; SEQ ID No 48957; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (II) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG0010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 89 AA;

Query Match 1.8%; Score 7; DB 22; Length 89;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 88 PSLGVSS 94  
|||  
Db 44 PSLGVSS 50

RESULT 14  
AAB32606  
ID AAB32606 standard; Protein; 96 AA.

XX AC AAB32606;

DT 25-JAN-2001 (first entry)

XX DE Eucalyptus grandis transcription factor protein sequence #64.

XX KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;  
XX poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;  
XX basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;  
XX homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;  
XX type 2 Cys2His2; CCAAT box element; MYB.

XX OS Eucalyptus grandis.

XX PN WO200053724-A2.

XX PD 14-SEP-2000.

XX PF 09-MAR-2000; 2000WO-US06112.

XX PR 11-MAR-1999; 99US-0266513.

XX PR 18-AUG-1999; 99US-0149485.

XX PA (GENE-) GENESIS RES & DEV CORP LTD.  
XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX PI Wood M, McGrath A, Shenk MA, Glenn M;

XX DR WPI; 2000-579369/54.

XX PT New isolated polynucleotide encoding a plant transcription factor for  
XX producing a plant e.g. a woody plant, preferably eucalyptus or pine,  
XX having modified gene expression or modified activity of a polypeptide

XX PS Claim 8; Page 225; 747pp; English.

XX CC The present invention relates to novel plant transcription factors from  
XX Eucalyptus grandis or Pinus radiata. The present sequence is one such  
XX transcription factor. The transcription factor may be used to produce a  
XX plant having modified gene expression such as a woody plant e.g. a  
XX eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or  
XX to modify the activity of a polypeptide in a plant. The transcription  
XX factors of the present invention are members from the following families  
XX of regulatory proteins: bZIP, bZIP family of G-box binding factors, basic  
XX helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeodomain  
XX zipper, LIM domain, AP2 and EREBs, zinc finger domains of type 2  
XX Cys2His2, CCAAT box elements and MYB.

SQ Sequence 96 AA;

Query Match 1.8%; Score 7; DB 21; Length 96;  
 Best Local Similarity 100.0%; Pred. No. 81;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 136 VARIEVL 142  
 |||||  
 66 VARIEVL 72

RESULT 15

ID AAB33167 standard; Protein; 96 AA.

AC AAB33167;

DT 25-JAN-2001 (first entry)

XX Eucalyptus grandis transcription factor protein sequence #356.

XX Plant; transcription factor; gene expression; eucalyptus; pine; acacia;

XX poplar; sweetgum; teak; mahogany; bZIP, G-box binding factor;

XX basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;

XX homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;

XX type 2 Cys2His2; CCAAT box element; MYB.

XX Eucalyptus grandis.

XX MO200053724-A2.

XX 14-SEP-2000.

XX 09-MAR-2000; 2000MO-US06112.

XX 11-MAR-1999; 99US-0266513.

XX 18-AUG-1999; 99US-0149485.

XX (GENE-) GENESIS RES &amp; DEV CORP LTD.

XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX Wood M, McGrath A, Shenk MA, Glenn M;

XX WPI; 2000-579369/54.

XX New isolated polynucleotide encoding a plant transcription factor for

XX producing a plant e.g. a woody plant, preferably eucalyptus or pine,

XX having modified gene expression or modified activity of a polypeptide

XX Claim 8; Page 663; 747pp; English.

XX The present invention relates to novel plant transcription factors from

XX Eucalyptus grandis or Pinus radiata. The present sequence is one such

XX plant having modified gene expression such as a woody plant e.g. a

XX eucalyptus, pine, acacia, poplar, sweetgum, teak, or mahogany species or

XX factors of the present invention are members from the following families

XX of regulatory proteins: bZIP, bZIP family of G-box binding factors, basic

XX helix-loop-helix zipper, homeotic/homeodomain/homeobox/MADS, homeodomain

XX Cys2His2, CCAAT box elements and MYB.

XX Sequence 96 AA;

XX Query Match 1.8%; Score 7; DB 21; Length 96;

XX Best Local Similarity 100.0%; Pred. No. 81;

XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX QY 136 VARIEVL 142

XX |||||

XX 66 VARIEVL 72

RESULT 16

ID AAG00628 standard; Protein; 96 AA.

AC AAG00628;

DT 06-OCT-2000 (first entry)

XX Human secreted protein, SEQ ID NO: 4709.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

XX gene therapy; chromosome mapping.

XX Homo sapiens.

XX EP1033401-A2.

XX 06-SEP-2000.

XX 21-FEB-2000; 2000EP-0200610.

XX 26-FEB-1999; 99US-0122487.

XX (GEST ) GENSET.

XX Dumas Milne Edwards J, Duclert A, Giordano J;

XX WPI; 2000-500381/45.

XX N-PSDB; AAC00634.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for

XX obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for

XX diagnostic, forensic, gene therapy and chromosome mapping procedures -

XX Claim 13; SEQ ID 4709; 71pp + CD-ROM; English.

XX The present sequence is a polypeptide encoded by one of a large number

XX of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs

XX were prepared from total human RNAs or polyA+ RNAs derived from 30

XX different tissues. EST sequences usually correspond mainly to the 3'

XX untranslated region (UTR) of the mRNA because they are often obtained

XX from oligo-dT primed cDNA libraries. Such ESTs are not well suited for

XX isolating cDNA sequences derived from the 5' ends of mRNAs and even in

XX those cases where longer cDNA sequences have been obtained, the full 5'

XX UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'

XX ends and can therefore be used to obtain full length cDNAs and genomic

XX DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and

XX chromosome mapping procedures. They are used to obtain upstream

XX regulatory sequences and to design expression and secretion vectors.

XX Sequence 96 AA;

XX Query Match 1.8%; Score 7; DB 21; Length 96;

XX Best Local Similarity 100.0%; Pred. No. 81;

XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX QY 199 SENGAVL 205

XX |||||

XX 62 SENGAVL 68

XX RESULT 17

XX ID AAY36092 standard; Protein; 113 AA.

XX AC AAY36092;

XX DT 13-SEP-1999 (first entry)

XX Extended human secreted protein sequence, SEQ ID NO. 477.

KW Secreted protein; human; cytokine; cellular proliferation; cell movement;  
 KW cellular differentiation; immune system regulator; anti-inflammatory;  
 KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;  
 KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;  
 KW genetic disease.

XX Homo sapiens.

XX WO9931236-A2.

XX 24-JUN-1999.

XX 17-DEC-1998; 98WO-IB02122.

XX 10-AUG-1998; 98US-0096116.

XX 17-DEC-1997; 97US-0069957.

XX 09-FEB-1998; 98US-0074121.

XX 13-APR-1998; 98US-0081563.

XX (GEST ) GENSET.

XX Bougueleret L, Duclert A, Dumas Milne Edwards J;

XX WPI; 1999-385906/32.

XX N-PSDB; AAX97776.

XX New isolated human secreted proteins

XX Claim 9; Page 411-412; 516pp; English.

XX This sequence is encoded by an extended human secreted protein coding  
 CC sequence of the invention. The secreted proteins can be used in treating  
 CC or controlling a variety of human conditions. The secreted proteins may  
 CC act as cytokines or may affect cellular proliferation or differentiation  
 CC or may act as immune system regulators, haematopoiesis regulators, tissue  
 CC growth regulators, regulators of reproductive hormones or cell movement  
 CC or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or  
 CC tumour inhibition activity. The DNAs can be used in forensic procedures  
 CC to identify individuals or in diagnostic procedures to identify  
 CC individuals having genetic diseases resulting from abnormal expression of  
 CC the genes corresponding to the extended cDNAs. They are also useful for  
 CC constructing a high resolution map of the human chromosomes. They can  
 CC also be used for gene therapy to control or treat genetic diseases.

XX Sequence 113 AA;

Query Match 1.8%; Score 7; DB 20; Length 113;

Best Local Similarity 100.0%; Pred. No. 95;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 199 SENGVL 205

|||||

Db 68 SENGVL 74

RESULT 18

AA52475

ID AA52475 standard; Protein; 114 AA.

XX AA52475;

XX 03-JUL-2002 (first entry)

XX Superoxide dismutase protein #4.

XX Selenoprotein; HIV; Ebola virus; cancer; immune system disorder.

XX Oryctolagus cuniculus.

XX US6303295-B1.

XX 16-OCT-2001.

XX

PP 12-JUL-1996; 96US-0679493.

XX 14-JUL-1995; 95US-001203P.

PR 01-SEP-1995; 95US-003112P.

XX (UYGE-) UNIV GEORGIA RES FOUND INC.

XX Taylor EW, Nadimpalli RG, Ramanathan CS;

XX WPI; 2002-024734/03.

XX New selenoprotein for use in detecting certain viruses, e.g. human

XX immunodeficiency virus (HIV) or Ebola, cancer and immune system

XX disorders -

XX Disclosure; Columns 77-88; 140pp; English.

XX The present invention relates to selenoproteins encoded in the genome of

XX a virus, where the coding sequence of the selenoprotein is genetically

XX engineered for expression in a nucleic acid construct. The invention also

XX discloses a method for identifying selenoprotein coding sequences, for

XX detecting certain viruses (e.g. HIV or Ebola), cancer and immune system

XX disorders. The present sequence was used to illustrate the invention.

XX Sequence 114 AA;

Query Match 1.8%; Score 7; DB 23; Length 114;

Best Local Similarity 100.0%; Pred. No. 96;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 358 PSATLDA 364

|||||

Db 29 PSATLDA 35

RESULT 19

AA51637

ID AA51637 standard; Protein; 126 AA.

XX AA51637;

XX 16-JUN-1999 (first entry)

XX Human 5' EST secreted protein SEQ ID NO:289.

XX Human; secreted protein; EST; expressed sequence tag; diagnosis;

XX forensic; gene therapy; chromosome mapping; signal peptide;

XX upstream regulatory sequence; cytokine activity; cell proliferation;

XX differentiation; haematopoiesis regulation; tissue growth regulation;

XX reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;

XX thrombolytic; anti-inflammatory; tumour inhibition.

XX Homo sapiens.

XX WO9906439-A2.

XX 11-FEB-1999.

XX 31-JUL-1998; 98WO-IB01233.

XX 01-AUG-1997; 97US-0904468.

XX (GEST ) GENSET.

XX Duclert A, Dumas Milne Edwards J, Lacroix B;

XX WPI; 1999-153700/13.

XX N-PSDB; AAX40355.

XX New nucleic acids encoding human secreted proteins - obtained from

XX cDNA libraries derived from liver, lung, large intestine, colon,

XX thyroid and pancreas tissue

XX

PS Claim 27; Page 370-371; 398pp; English.  
 CC AA40251 to AA40397 represent 5' expressed sequence tags (ESTs) for  
 CC human secreted proteins and encode the proteins given in AA11533 to  
 CC AA11679, respectively. The proteins given represent the signal peptide  
 CC and an N-terminal fragment of a secreted protein. The nucleic acid  
 CC sequences can be used for producing secreted human gene products. They  
 CC can also be used to develop products for diagnosis and therapy. The  
 CC proteins obtained may have cytokine activity, cell  
 CC proliferation/differentiation activity, haematopoiesis regulating  
 CC activity, tissue growth regulating activity, reproductive hormone  
 CC regulating activity, chemotactic/chemokinetic activity, haemostatic and  
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, tumour inhibition activity or other activities. The products  
 CC can be used in forensic, gene therapy and chromosome mapping procedures.  
 CC The sequences can also be used for obtaining corresponding promoter  
 CC sequences. The nucleic acids encoding the signal peptide can be used for  
 CC directing extracellular secretion of a polypeptide or the insertion of a  
 CC polypeptide into a membrane, or importing a polypeptide into a cell.  
 CC  
 SQ Sequence 126 AA;

Query Match 1.8%; Score 7; DB 20; Length 126;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 323 INTNAVY 329  
 |||||  
 Db 55 INTNAVY 61

RESULT 20  
 ABG20115  
 ID ABG20115 standard; Protein; 145 AA.

AC ABG20115;  
 DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #20106.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSB-) HYSBQ INC.

PI Drmanac RT, Liu C, Tang YT.

DR WPI; 2001-639362/73.

DR N-PSDB; AAS84302.

PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostic, forensic, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -

PS Claim 20; SEQ ID No 50474; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations in  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 145 AA;

Query Match 1.8%; Score 7; DB 22; Length 145;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LSLPLUS 16  
 |||||  
 Db 3 LSLPLUS 9

RESULT 21  
 AAM20959  
 ID AAM20959 standard; Protein; 156 AA.

AC AAM20959;

DT 21-JUL-1997 (first entry)

DE H. pylori cytoplasmic protein hsp13939orf13.

KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;  
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.

OS Helicobacter pylori.

PN WO9640893-A1.

PD 19-DEC-1996.

PF 06-JUN-1996; 96WO-US09122.

PR 01-APR-1996; 96US-0630405.

PR 07-JUN-1995; 95US-0487032.

PA (ASTR ) ASTRA AB.

PI Berglindh OT, Smith D, Mellgaard BL.

DR WPI; 1997-052306/05.

DR N-PSDB; AAT68212.

PT Helicobacter pylori nucleic acid sequences and related  
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
 PT infection, and to detect Helicobacter

PS Claim 61; Page 1350; 1481pp; English.

CC The present sequence represents a Helicobacter pylori cytoplasmic  
 CC protein that may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori life cycle activators or inhibitors.  
 CC useful as potential H. pylori life cycle activators or inhibitors.  
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
 CC overlapping contigs generated by mechanically shearing the bacterial

CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,  
 CC and the predicted coding regions defined by computer evaluation. To  
 CC identify likely H. pylori antigens for vaccine development, the amino  
 CC acid sequences predicted from various ORF were analysed for significant  
 CC homology to other known or exported membrane proteins. Having identified  
 CC and determined the sequences of interest, particular regions can be  
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
 CC production, e.g. in E. coli hosts.

XX SQ Sequence 156 AA;  
 Query Match 1.8%; Score 7; DB 18; Length 156;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 LSLPLLS 16  
 |||||  
 Db 24 LSLPLLS 30

RESULT 22  
 AAW20518  
 ID AAW20518 standard; Protein; 164 AA.  
 XX  
 AC AAW20518;  
 XX  
 DT 14-JUL-1997 (first entry)  
 XX  
 DE H. pylori cytoplasmic protein 4766691.aa.  
 XX  
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;  
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;  
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.  
 XX  
 OS Helicobacter pylori.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 12  
 FT /label= unknown  
 FT /note= "encoded by ARG"  
 FT Misc-difference 100  
 FT /label= unknown  
 FT /note= "encoded by GRC"  
 XX  
 FN WO9640893-A1.  
 XX  
 PD 19-DEC-1996.  
 XX  
 PF 06-JUN-1996; 96WO-US09122.  
 XX  
 PR 01-APR-1996; 96US-0630405.  
 PR 07-JUN-1995; 95US-0487032.  
 XX  
 PA (ASTR ) ASTRA AB.  
 XX  
 FI Berglindh OT, Smith D, Mellgaard BL;  
 XX  
 DR WPI; 1997-052306/05.  
 DR N-PSDB; AAT67674.  
 XX  
 XX Helicobacter pylori nucleic acid sequences and related  
 FT polypeptide(s) - useful for vaccines to treat or prevent H. pylori  
 FT infection, and to detect Helicobacter  
 XX  
 PS Claim 61; Page 679; 1481pp; English.  
 XX  
 CC This sequence is a H. pylori cytoplasmic protein.  
 CC The protein may be used in a vaccine to prevent or treat H. pylori  
 CC infection or to identify H. pylori polypeptide binding compounds,  
 CC useful as potential H. pylori life cycle activators or inhibitors.  
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from  
 CC overlapping contigs generated by mechanically shearing the bacterial  
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,

CC and the predicted coding regions defined by computer evaluation. To  
 CC identify likely H. pylori antigens for vaccine development, the amino  
 CC acid sequences predicted from various ORF were analysed for significant  
 CC homology to other known or exported membrane proteins. Having identified  
 CC and determined the sequences of interest, particular regions can be  
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide  
 CC production, e.g. in E. coli hosts.

XX SQ Sequence 164 AA;  
 Query Match 1.8%; Score 7; DB 18; Length 164;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 10 LSLPLLS 16  
 |||||  
 Db 21 LSLPLLS 27

RESULT 23  
 ABB58319  
 ID ABB58319 standard; Protein; 183 AA.  
 XX  
 AC ABB58319;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 1749.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 FI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL02422.  
 XX  
 FT New isolated nucleic acid detection reagent for detecting 1000 or more  
 FT genes from Drosophila and for elucidating cell signalling and cell-cell  
 FT interactions -  
 XX  
 PS Disclosure; SEQ ID NO 1749; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABLI16176-ABLI30511), expressed DNA  
 CC sequences (ABLI01840-ABLI16175) and the encoded proteins  
 CC (ABBS7737-ABBS72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 183 AA;  
 Query Match 1.8%; Score 7; DB 22; Length 183;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 164 SDLEEQ 170  
 PN |||||  
 XX  
 Db 65 SDLEEQ 71

RESULT 24

AAW62792  
 ID AAW62792 standard; Protein; 187 AA.

AC AAW62792;

DT 25-SEP-1998 (first entry)

DE Tryparedoxin peroxidase amino acid sequence.

KM Tryparedoxin peroxidase; inhibitor; trypanocidal agent.

XX Cricitidia fasciculata.

PN WO9826052-A1.

PD 18-JUN-1998.

PF 12-SEP-1997; 97WO-EP04990.

PR 12-DEC-1996; 96EP-0120016.

PA (FLOH/) FLOHE L.

PI Kallaz H, Montemartini M, Nogeceke E;

DR WPI; 1998-348515/30.

DR N-PSDB; AAW42338.

PT New tryparedoxin peroxidase - useful for, e.g. identification of

PS trypanocidal agents

XX Claim 7; Page 21; 39pp; English.

CC The present sequence represents a tryparedoxin peroxidase from Cricitidia  
 CC fasciculata. This protein is able to transfer reducing equivalents of  
 CC trypanthione (T(SH)2), or its disulphide (T(S2)) to hydrogen peroxide  
 CC and/or an alkyl hydroperoxide, via a mediator such as tryparedoxin that  
 CC mediates between T(SH)2/T(S2) tryparedoxin peroxidase. Tryparedoxin  
 CC peroxidase, is used to test for and isolate specific inhibitors of  
 CC itself. Such inhibitors are potentially useful as trypanocidal agents.

XX Sequence 187 AA;

Query Match 1.8%; Score 7; DB 19; Length 187;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 371 ARAYGVL 377  
 PN |||||  
 XX  
 Db 113 ARAYGVL 119

RESULT 25

AAW624217  
 ID AAW624217 standard; Protein; 195 AA.

AC AAW624217;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 27805.

KM Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 XX termination sequence.

OS Arabidopsis thaliana.

XX EPI03405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

```
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 22-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 27-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 27-AUG-1999; 99US-0151303.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.

PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 1.8%; Score 7; DB 21; Length 195;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 348 ALPILQS 354
Db 74 ALPILQS 80
|||||
|||||

RESULT 26
AAR11535
ID AAR11535 standard; Protein; 209 AA.
XX
AC AAR11535;
XX
DT 14-JUN-1991 (first entry)
XX
DE EC-SOD mutant, T209.
XX
KW Super oxide dismutase; heparin; ischaemia; inflammation; truncation;
XX Homo sapiens.
XX WO9104315-A.
XX
PD 04-APR-1991.
XX
PF 17-SEP-1990; 90WO-DK00238.
XX
PR 15-SEP-1989; 89DK-0004557.
```





Oy 122 RDISQL 128  
 Db 168 RDISQL 174

## RESULT 29

AAR11537  
 ID AAR11537 standard; Protein; 215 AA.

XX AC AAR11537;  
 DT 14-JUN-1991 (first entry)

XX DE EC-SOD mutant, T215.

XX KW Super oxide dismutase; heparin; ischaemia; inflammation; truncation;  
 XX OS Homo sapiens.

XX PN WO9104315-A.

XX PD 04-APR-1991.

XX PF 17-SEP-1990; 90WO-DK00238.

XX PR 15-SEP-1989; 89DK-0004557.

XX PA (SYMB-) SYMBICOM AB.

XX PI Marklund S, Edlund T;

XX DR WPI; 1991-117506/16.

XX DR N-PSDB; AAQ11356.

XX PT Extracellular super oxide dismutase variants (EC-SOD) - for  
 PT treatment of conditions associated with super oxide radicals such  
 PT as ischaemia and inflammatory diseases.

XX PS Claim 9; Fig 1A; 125pp; English.

XX CC The sequence is prepd. by mutagenesis of human EC-SOD type C cDNA.  
 CC Codon -gag (encoding residue no. 216) is replaced with a STOP  
 CC codon resulting in the expression of the truncated EC-SOD variant  
 CC T215 by the plasmid pPST215. The variant polypeptide has the  
 CC enzymatic activity of native extracellular EC-SOD but no heparin  
 CC binding activity (EC-SOD class A), improving migration to part-  
 CC icular sites in vivo where activity is required and useful when a  
 CC high activity in plasma or interstitial fluid is important.  
 CC See also AAR11535-R11544.

XX SQ Sequence 215 AA;

Query Match 1.8%; Score 7; DB 12; Length 215;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 358 PSATLDA 364  
 Db 49 PSATLDA 55

## RESULT 30

AAR11538  
 ID AAR11538 standard; Protein; 216 AA.

XX AC AAR11538;

XX DT 14-JUN-1991 (first entry)

XX DE EC-SOD mutant, T216.

XX KW Super oxide dismutase; heparin; ischaemia; inflammation; truncation;

OS Homo sapiens.

XX PN WO9104315-A.

XX PD 04-APR-1991.

XX PF 17-SEP-1990; 90WO-DK00238.

XX PR 15-SEP-1989; 89DK-0004557.

XX PA (SYMB-) SYMBICOM AB.

XX PI Marklund S, Edlund T;

XX DR WPI; 1991-117506/16.

XX DR N-PSDB; AAQ11357.

XX PT Extracellular super oxide dismutase variants (EC-SOD) - for  
 PT treatment of conditions associated with super oxide radicals such  
 PT as ischaemia and inflammatory diseases.

XX PS Claim 8; Fig 1A; 125pp; English.

XX CC The sequence is prepd. by mutagenesis of human EC-SOD type C cDNA.  
 CC Codon -agc (encoding residue no. 217) is replaced with a STOP  
 CC codon resulting in the expression of the truncated EC-SOD variant  
 CC T216 by the plasmid pPST216. The variant polypeptide has the  
 CC enzymatic activity of native extracellular EC-SOD but slightly  
 CC reduced affinity for heparin (it is still EC-SOD class C).

XX CC improving migration to particular sites in vivo where activity is  
 CC required and useful when a high activity in plasma or interstitial  
 CC fluid is important.

XX CC See also AAR11535-R11544.

XX SQ Sequence 216 AA;

Query Match 1.8%; Score 7; DB 12; Length 216;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 358 PSATLDA 364  
 Db 49 PSATLDA 55

## RESULT 31

AAR11544

ID AAR11544 standard; Protein; 216 AA.

XX AC AAR11544;

XX DT 14-JUN-1991 (first entry)

XX DE EC-SOD mutant, SAT216.

XX KW Super oxide dismutase; heparin; ischaemia; inflammation; truncation;

XX OS Homo sapiens.

XX PN WO9104315-A.

XX PD 04-APR-1991.

XX PF 17-SEP-1990; 90WO-DK00238.

XX PR 15-SEP-1989; 89DK-0004557.

XX PA (SYMB-) SYMBICOM AB.

XX PI Marklund S, Edlund T;

XX DR WPI; 1991-117506/16.

XX DR N-PSDB; AAQ11363.

```
XX Extracellular super oxide dismutase variants (EC-SOD) - for
PT treatment of conditions associated with super oxide radicals such
PT as ischaemia and inflammatory diseases.
XX
PS Claim 9; Fig 1C; 125pp; English.
XX
CC The sequence is prep'd. by mutagenesis of human EC-SOD type C cDNA.
CC Codons -gag agc- (encoding residues no. 214-215; Glu-Ser) are
CC replaced with -ggc tga- encoding an Ala residue followed by a STOP
CC codon. This results in the expression of the variant SAT216, by
CC plasmid pPSA216 which has a higher net positive charge than the
CC native polypeptide. The variant has the enzymatic activity of native
CC polypeptide but has higher affinity for heparin. This is useful where
CC for eg organ transplantations.
CC See also AAR1535-R11543.
CC
SQ Sequence 216 AA;

Query Match 1.8%; Score 7; DB 12; Length 216;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 PSATLDA 364
DB 49 PSATLDA 55

RESULT 32
AAR1540
ID AAR1540 standard; Protein; 222 AA.
XX
AC AAR1540;
XX
DT 14-JUN-1991 (first entry)
XX
DE EC-SOD mutant, SA219.
XX
KM Super oxide dismutase; heparin; ischaemia; inflammation; truncation;
XX
OS Homo sapiens.
XX
PN WO9104315-A.
XX
PD 04-APR-1991.
XX
PF 17-SEP-1990; 90WO-DK00238.
XX
PR 15-SEP-1989; 89DK-0004557.
XX
PA (SYMB-) SYMBICOM AB.
XX
PI Marklund S, Edlund T;
XX
WPI; 1991-117506/16.
DR N-PSDB; AAQ11359.
XX
PT Extracellular super oxide dismutase variants (EC-SOD) - for
PT treatment of conditions associated with super oxide radicals such
PT as ischaemia and inflammatory diseases.
XX
PS Disclosure; Fig 1B; 125pp; English.
XX
CC The sequence is prep'd. by mutagenesis of human EC-SOD type C cDNA.
CC Codon -tgc- (encoding residue no. 219; Cys) is replaced with -ggc-
CC encoding an Ala residue, resulting in the loss of a disulphide
CC bridge which maintains the structure of the C-terminal. This has
CC no effect on the heparin affinity of the variant as expressed by
CC the plasmid pPSA219 which also has the enzymatic activity of native
CC extracellular EC-SOD.
CC See also AAR1535-R11544.
CC
SQ Sequence 222 AA;
```

```
Query Match 1.8%; Score 7; DB 12; Length 222;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 PSATLDA 364
DB 49 PSATLDA 55

RESULT 33
AAR1539
ID AAR1539 standard; Protein; 222 AA.
XX
AC AAR1539;
XX
DT 14-JUN-1991 (first entry)
XX
DE EC-SOD mutant, SA216.
XX
KM Super oxide dismutase; heparin; ischaemia; inflammation; truncation;
XX
OS Homo sapiens.
XX
PN WO9104315-A.
XX
PD 04-APR-1991.
XX
PF 17-SEP-1990; 90WO-DK00238.
XX
PR 15-SEP-1989; 89DK-0004557.
XX
PA (SYMB-) SYMBICOM AB.
XX
PI Marklund S, Edlund T;
XX
WPI; 1991-117506/16.
DR N-PSDB; AAQ11358.
XX
PT Extracellular super oxide dismutase variant (EC-SOD) - for
PT treatment of conditions associated with super oxide radicals such
PT as ischaemia and inflammatory diseases.
XX
PS Claim 9; Fig 1B; 125pp; English.
XX
CC The sequence is prep'd. by mutagenesis of human EC-SOD type C cDNA.
CC Codon -gag- (encoding residue no. 216; Glu) is replaced with -ggc-
CC encoding an Ala residue, resulting in a greater net positive
CC charge. The variant, SA216, expressed by the plasmid pPSA216 has
CC the enzymatic activity of native extracellular EC-SOD and a greater
CC affinity for heparin than the native EC-SOD class C, useful for
CC eg organ transplantation.
CC See also AAR1535-R11544.
CC
SQ Sequence 222 AA;

Query Match 1.8%; Score 7; DB 12; Length 222;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 PSATLDA 364
DB 49 PSATLDA 55

RESULT 34
AAR1541
ID AAR1541 standard; Protein; 222 AA.
XX
AC AAR1541;
XX
DT 14-JUN-1991 (first entry)
XX
```

```

DE EC-SOD mutant, SA220.
XX
KW Super oxide dismutase; heparin; ischaemia; inflammation; truncation;
XX
OS Homo sapiens.
XX
PN WO9104315-A.
XX
PD 04-APR-1991.
XX
XX
PF 17-SEP-1990; 90WO-DK00238.
XX
PR 15-SEP-1989; 89DK-0004557.
XX
PA (SYMB-) SYMBICOM AB.
XX
PI Marklund S, Edlund T;
XX
DR WPI; 1991-117506/16.
DR N-PSDB; AAQ11360.
XX
XX Extracellular super oxide dismutase variants (EC-SOD) - for
PT treatment of conditions associated with super oxide radicals such
PT as ischaemia and inflammatory diseases.
XX
PS Claim 9; Fig 1B; 125pp; English.
XX
XX The sequence is prep'd. by mutagenesis of human EC-SOD type C cDNA.
CC Codon -aag- (encoding residue no. 220; Arg) is replaced with -gcg-
CC encoding an Ala residue, resulting in a decrease in the net
CC positive charge. The variant, SA220, expressed by the plasmid
CC pPSA220, has the enzymatic activity of native polypeptide but has
CC a slightly reduced affinity for heparin. This is useful where
CC migration to a particular site in vivo is required or high levels of
CC circulating enzyme are needed.
CC See also AAR11535-R11544.
XX
XX Sequence 222 AA;
SQ
Query Match 1.8%; Score 7; DB 12; Length 222;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 358 PSATLDA 364
Db 49 PSATLDA 55

RESULT 35
AAR11542
ID AAR11542 standard; Protein; 222 AA.
XX
AC AAR11542;
XX
DT 14-JUN-1991 (first entry)
XX
DE EC-SOD mutant, G1.
XX
KW Super oxide dismutase; heparin; ischaemia; inflammation; truncation;
XX
OS Homo sapiens.
XX
PN WO9104315-A.
XX
PD 04-APR-1991.
XX
XX
PF 17-SEP-1990; 90WO-DK00238.
XX
PR 15-SEP-1989; 89DK-0004557.
XX
PA (SYMB-) SYMBICOM AB.
XX
PI Marklund S, Edlund T;

```

```

XX
DR WPI; 1991-117506/16.
DR N-PSDB; AAQ11361.
XX
XX Extracellular super oxide dismutase variants (EC-SOD) - for
PT treatment of conditions associated with super oxide radicals such
PT as ischaemia and inflammatory diseases.
XX
PS Claim 12; Fig 1B; 125pp; English.
XX
XX The sequence is prep'd. by mutagenesis of human EC-SOD type C cDNA.
CC Codon -aac- (encoding residue no. 89; Asn) is replaced with -caa-
CC encoding a Gln residue, resulting in a glycosylation-free variant.
CC The variant, G1, expressed by the plasmid pPSG1, has the enzymatic
CC activity of native polypeptide but has an increased affinity for
CC heparin. This may be due to the fact that the CHO substituent
CC carries a considerable negative charge which may repel the enzyme
CC from the strongly negatively charged heparin. The increased
CC affinity is useful eg for organ transplantation.
CC See also AAR11535-R11544.
XX
XX Sequence 222 AA;
SQ
Query Match 1.8%; Score 7; DB 12; Length 222;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 358 PSATLDA 364
Db 49 PSATLDA 55

RESULT 36
AAR64910
ID AAR64910 standard; Protein; 222 AA.
XX
AC AAR64910;
XX
DT 11-AUG-1995 (first entry)
XX
DE Human extracellular superoxide-dismutase (SOD) partial sequence.
XX
KW Superoxide-dismutase; SOD; enzyme; transgenic mammal; mammary gland;
KW mamma; milk; ischaemia; antiinflammatory; cosmetics; gene transfer.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Modified-site 89
FT Active-site /note= "glycosylation site"
FT Misc-difference 97..193
FT Misc-difference 1..96 /note= "involved in oligomer formation"
XX
XX
PN WO9500637-A.
XX
PD 05-JAN-1995.
XX
XX
PF 24-JUN-1994; 94WO-IB00181.
XX
PR 24-JUN-1993; 93DK-0000753.
XX
PA (SYMB-) SYMBICOM AB.
XX
PI Hansson L;
XX
DR WPI; 1995-052074/07.
DR N-PSDB; AAQ76316.
XX
XX Prodn. of recombinant human extracellular superoxide dismutase -
PT using transgenic mammals to express the polypeptide and useful
PT variants combined as hybrids with a milk protein in the mammary
PT gland

```

XX Claim 1; Page 69; 103pp; English.  
PS  
CC This protein may be expressed (using plasmid pSI72) in the mammary  
XX gland of an adult female of a non-human mammal harboring a SOD  
CC gene-milk protein gene regulatory element hybrid gene. The protein  
CC may be used for the diagnosis, prophylaxis and treatment of  
CC diseases or disorders related to the presence of superoxide  
CC radicals. It may also be used essentially for the same  
CC applications as CuZn SOD. The protein may also be useful in  
CC cosmetics, e.g. to prevent skin ageing and protect against  
CC radiation damage. Modified proteins may be produced from this  
CC sequence which have increased/decreased heparin affinity.  
XX  
SQ Sequence 222 AA;  
Query Match 1.8%; Score 7; DB 16; Length 222;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 358 PSATLDA 364  
DB 49 PSATLDA 55  
RESULT 37  
AAR25161  
ID AAR25161 standard; Protein; 223 AA.  
XX  
AC AAR25161;  
XX  
DT 14-DEC-1992 (first entry)  
XX  
DE EC-SOD.  
XX  
KM Ischaemic cardiac disorders; myocardial infarction; collagen diseases;  
XX radiation disorders.  
XX  
OS Synthetic.  
XX  
PN EP492447-A.  
XX  
PD 01-JUL-1992.  
XX  
PF 19-DEC-1991; 9LEP-0121802.  
XX  
PR 28-DEC-1990; 90UP-0416786.  
XX  
PA (TOXN ) TOYO JOZO KK.  
XX  
PI Akashi H, Katayama K, Mizoguchi J, Nagoshi H, Otsu T;  
PI Yamashita T;  
PI  
DR WPI; 1992-218663/27.  
DR N-PSDB; AAQ26057.  
XX  
XX  
PT Prep. of active purified extracellular superoxide dismutase - by  
PT forming solubilised monomer from insol. recombinant aggregates, -  
PT refolding and purifying, used for treating Behcet disease, etc.  
XX  
PS Disclosure; Page 8-9; 20pp; English.  
XX  
XX The sequence given is enzymatically active purified extracellular  
XX superoxide dismutase (EC-SOD). It is produced from insol. inactive  
XX aggregates containing recombinant EC-SOD by treating the insol.  
XX active aggregates under a chaotropic environment or reducing  
XX conditions until the EC-SOD, in the form of a solubilised monomer is  
XX formed and recovered, refolding the monomeric EC-SOD recovered into  
XX an active EC-SOD and purifying the refolded active EC-SOD. EC-SOD  
XX is a glycoprotein which consists of four subunits and has a sugar  
XX chain of approx. 135kD. It has four copper and four zinc ions per  
XX molecule. The EC-SOD produced in this invention has a high  
XX enzymatic activity of 3000 unit/mg of protein or more. This EC-SOD

CC can be used in the treatment of Behcet's disease, ischaemic cardiac  
CC disorders such as myocardial infarction, collagen diseases or  
CC radiation disorders.  
XX  
SQ Sequence 223 AA;  
Query Match 1.8%; Score 7; DB 13; Length 223;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 358 PSATLDA 364  
DB 50 PSATLDA 56  
RESULT 38  
AAG24251  
ID AAG24251 standard; Protein; 226 AA.  
XX  
AC AAG24251;  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 27852.  
XX  
KM Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134370.  
PR 19-MAY-1999; 99US-0134768.  
PR 20-MAY-1999; 99US-0134941.  
PR 21-MAY-1999; 99US-0135124.  
PR 24-MAY-1999; 99US-0135353.  
PR 25-MAY-1999; 99US-0135629.  
PR 27-MAY-1999; 99US-0136021.  
PR 28-MAY-1999; 99US-0136392.  
PR 01-JUN-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139859.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 24-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140981.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 11-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 24-SEP-1999; 99US-0155486.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160788.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match

1.8%; Score 7; DB 21; Length 226;



KW reverse transcription polymerase chain reaction; SSCP analysis.

```
XX OS Homo sapiens.
XX PH Location/Qualifiers
FT Peptide 1..24
FT Protein /label= signal peptide
FT Protein 25..240
FT Protein /label= mature_EC-SOD
XX WO9419493-A.
XX 01-SEP-1994.
XX 28-FEB-1994; 94WO-US02089.
XX 26-FEB-1993; 93US-0023980.
XX (GEHO ) GEN HOSPITAL CORP.
XX (MASI ) MASSACHUSETTS INST TECHNOLOGY.
XX Brown R, Horvitz HR, Rosen DR;
XX WPI; 1994-294353/36.
XX N-PSDB; AAQ67475.
XX Diagnosis, treatment and prevention of diseases of cell death -
XX e.g. amyotrophic lateral sclerosis, which are the result of e.g.
XX decreased SOD activity
XX Claim 22; Fig 4; 94pp; English.
XX The presence of a mutation in a gene encoding a superoxide
XX dismutase (SOD1, SOD2 or SOD3) indicates an increased likelihood of
XX developing a cell death disease, specifically a neurodegenerative
XX disease. The use of SOD polypeptides to treat amyotrophic lateral
XX sclerosis and diseases involving a deleterious mutation in the
XX glutathione peroxidase-encoding gene, the catalase-encoding gene
XX and the nitric oxide-encoding gene is claimed. The specification
XX includes the sequences of human SOD1, hSOD2 and hSOD3 (AAR61336-
XX AAR61338, respectively).
XX Sequence 240 AA;
XX Query Match 1.8%; Score 7; DB 15; Length 240;
XX Best Local Similarity 100.0%; Pred. No. 2e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 358 PSATLDA 364
Db 67 PSATLDA 73
RESULT 42
AAR72413
ID AAR72413 standard; Protein; 240 AA.
XX AC AAR72413;
XX 30-OCT-1995 (first entry)
XX Human EC-SOD.
XX EC-SOD; superoxide dismutase.
XX Homo sapiens.
XX Key Location/Qualifiers
XX Peptide 1..18
XX Protein /label= signal peptide
XX Protein 19..240
XX Protein /label= mature protein
```

```
PN WO9510185-A.
XX 20-APR-1995.
XX 13-OCT-1994; 94WO-US11558.
XX 15-OCT-1993; 93US-0136207.
XX 20-SEP-1994; 94US-0136207.
XX (UYAL-) UNIV ALABAMA.
XX (UYDU-) UNIV DUKE.
XX Crapo JD, Day BJ, Foiz RJ, Freeman BA, Fridovich I;
XX Cury T, Polzrj;
XX WPI; 1995-161483/21.
XX N-PSDB; AAQ88057.
XX Super:oxide dismutase mimetic(s) comprising a nitrogen-containing
XX macrocyclic moiety - useful in treatment of inflammation,
XX musculation disorders, atherosclerosis or gout.
XX Disclosure; Figure 24; 136pp; English.
XX Clone 7 from a human adult leukocyte genomic library was analysed.
XX The analysis revealed that human EC-SOD contains three exons and
XX two introns. The identification of the signal peptide cleavage
XX site is consistent with computer algorithms which predict the
XX site of eukaryotic signal peptide cleavage.
XX Sequence 240 AA;
XX Query Match 1.8%; Score 7; DB 16; Length 240;
XX Best Local Similarity 100.0%; Pred. No. 2e+02;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 358 PSATLDA 364
Db 67 PSATLDA 73
RESULT 43
AAW32599
ID AAW32599 standard; Protein; 240 AA.
XX AC AAW32599;
XX 06-FEB-1998 (first entry)
XX Human EC-SOD protein.
XX Oxidant scavenger; extracellular matrix targeting moiety;
XX cell surface matrix; nitrogen containing macrocyclic moiety;
XX inflammatory condition; aberrant smooth muscle function;
XX ischaemia reperfusion injury; myocardial infarction; stroke;
XX acute head trauma; organ reperfusion; bowel ischaemia;
XX pulmonary infarction; glaucoma; EC-SOD; superoxide dismutase;
XX tetrameric glycosylated copper and zinc containing enzyme.
XX Homo sapiens.
XX WO9640223-A1.
XX 19-DEC-1996.
XX 07-JUN-1996; 96WO-US10497.
XX 11-MAR-1996; 96US-0613418.
XX 07-JUN-1995; 95US-0476866.
XX (TROV/) TROVA M P.
XX (UYAL-) UNIV ALABAMA.
XX (UYDU-) UNIV DUKE.
```



XX Batinic-Haberle I, Crapo JD, Day BJ, Folz RJ, Freeman BA;  
 PI Fridovich I, Oury T, Trova MP;  
 XX  
 XX WPI: 1997-077220/07.  
 DR N-PSDB; AAT92317.  
 XX  
 PT New porphyrin-type oxidant scavengers - used for protecting against  
 PT oxidants and for modulating biological processes involving oxidants.  
 XX  
 PS Claim 24; Fig 24; 195pp; English.

CC Oxidant scavengers have been developed comprising a nitrogen containing  
 CC macrocyclic moiety and a cell surface or extracellular matrix targeting  
 CC moiety, or their salts. The present sequence represents the human  
 CC EC-SOD, a tetrameric glycosylated copper and zinc containing superoxide  
 CC dismutase enzyme. The oxidant scavengers can be used for protecting  
 CC against the deleterious effects of oxidants and for modulating  
 CC biological processes involving oxidants. They can be used for e.g.  
 CC treating inflammatory conditions, treating disorders resulting from  
 CC reperfusion injuries associated with myocardial infarction, stroke,  
 CC acute head trauma, organ reperfusion following transplantation, bowel  
 CC ischaemia, pulmonary infarction, surgical occlusion of blood flow, and  
 CC soft tissue injury. They can further be used to protect against damage  
 CC to the eye due to sunlight (and to the skin) as well as glaucoma, and  
 CC macular degeneration of the eye. Diseases of the bone are also amenable  
 CC to treatment with the compounds, and connective tissue disorders  
 CC associated with defects in collagen synthesis or degradation can be  
 CC treated with the compounds. In the oxidant scavengers, substituents can  
 CC be selected to render them resistant to degradation by haemoxygenase and  
 CC also so that they do not interfere with normal porphyrin metabolism, can  
 CC pass through cell membranes and bind to cell surface or extracellular  
 CC matrix elements.

SO Sequence 240 AA;

Query Match 1.8%; Score 7; DB 18; Length 240;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 PSATLDA 364  
 |||||  
 DB 67 PSATLDA 73

RESULT 44  
 AAW82447  
 ID AAW82447 standard; Protein; 240 AA.  
 XX  
 AC AAW82447;  
 XX  
 DT 24-FEB-1999 (first entry)  
 XX  
 DE Human eCSOD protein.  
 XX  
 XX SOD1; SOD2; SOD3; Cu/Zn: superoxide dismutase; mitochondrial; treatment;  
 KM extracellular; neurodegenerative disease; amyotrophic lateral sclerosis;  
 KM familial; ALS.  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX US5849290-A.  
 PN  
 XX  
 PD 15-DEC-1998.  
 XX  
 PF 07-JUN-1995; 95US-0486953.  
 XX  
 PR 28-FEB-1994; 94US-0204052.  
 PR 26-FEB-1993; 93US-0023980.  
 PR 07-JUN-1995; 95US-0486953.  
 XX  
 PA (GEHO ) GEN HOSPITAL CORP.

PA (MAST ) MASSACHUSETTS INST TECHNOLOGY.  
 XX  
 XX Brown R, Horvitz HR, Rosen DR;  
 XX  
 XX WPI: 1999-069657/06.  
 DR N-PSDB; AAV73825.  
 XX  
 XX Treatment of neurodegenerative disease - by administering  
 PT super-oxide dismutase  
 XX  
 PS Disclosure; Fig 4; 53pp; English.

CC This sequence is a novel human extracellular SOD (superoxide dismutase)  
 CC protein encoded by the SOD3 gene. This protein can be used in a method  
 CC for treating a neurodegenerative disease particularly familial  
 CC amyotrophic lateral sclerosis (ALS).

SO Sequence 240 AA;

Query Match 1.8%; Score 7; DB 20; Length 240;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 PSATLDA 364  
 |||||  
 DB 67 PSATLDA 73

RESULT 45  
 AAY94782  
 ID AAY94782 standard; Protein; 240 AA.  
 XX  
 AC AAY94782;  
 XX  
 DT 12-FEB-2001 (first entry)  
 XX  
 DE Human superoxide dismutase (EC-SOD) protein sequence.

KM Porphine; metal complex oxidant scavenger; inhibitor; EC-SOD; gout;  
 KM superoxide dismutase; myocardial infarction; stroke; acute head trauma;  
 KM organ reperfusion; bowel ischaemia; pulmonary infarction; glaucoma;  
 KM skeletal muscle reperfusion injury; central nervous system disease; AIDS;  
 KM dementia; stroke; amyotrophic lateral sclerosis; Parkinson's disease;  
 KM Huntington's disease; neurological disorder; arthritis; hypertension;  
 KM arteriosclerosis; oedema; septic shock; pulmonary hypertension; asthma;  
 KM impotence; infertility; endometriosis; diabetes; pneumonia; human;  
 KM cystic fibrosis; sinusitis; autoimmune disease.

OS  
 XX  
 XX Homo sapiens.  
 OS  
 XX  
 XX Key Location/Qualifiers  
 FH Peptide 1..18  
 FT /note= "Putative signal peptide"  
 FT Protein 19..240  
 FT /label= "EC-SOD"  
 FT /note= "Superoxide dismutase"  
 XX  
 PN US6127356-A.  
 XX  
 PD 03-OCT-2000.  
 PD  
 PF 07-JUN-1996; 96US-0663028.  
 XX  
 PR 15-OCT-1993; 93US-0136207.  
 PR 13-OCT-1994; 94US-0322766.  
 PR 07-JUN-1995; 95US-0476866.  
 PR 11-MAR-1996; 96US-0613418.  
 XX  
 XX (UYDG-) UNIV DUKE.  
 PA  
 PI Crapo JD, Fridovich I, Oury T, Folz RJ, Trova MP, Freeman BA;  
 PI Batinic-Haberle I, Day BJ;  
 XX

DR WPI: 2000-664150/64.  
 DR N-PSDB; AAA28294.  
 XX  
 PT New metal complexes of methine substituted porphines useful as  
 PT catalytic oxygen scavengers -  
 XX  
 PS Disclosure; Fig 24; 97pp; English.  
 XX  
 CC This invention relates to porphines and their metal complex oxidant  
 CC scavengers, where the metal is manganese, copper or iron. The porphines  
 CC exhibit cardiant; cerebroprotective; vasotropic; ophthalmological;  
 CC antiparkinsonian; nootropic; anticonvulsant; cytostatic; gynecological;  
 CC antiarteriosclerotic; antiinflammatory; antibacterial; immunosuppressive;  
 CC hypotensive; antidiabetic; antitumor; antiaesthetic; and virucide  
 CC activity. The porphines act as superoxide dismutase (SOD) inhibitors and  
 CC are used as catalytic scavengers of reactive oxygen species to protect  
 CC against ischaemia reperfusion injuries associated with myocardial  
 CC infarction, stroke, acute head trauma, organ reperfusion following  
 CC transplantation, bowel ischaemia, pulmonary infarction, surgical  
 CC occlusion of blood flow, soft tissue injury, skeletal muscle reperfusion  
 CC injuries, glaucoma, macular degeneration of the eye, diseases of the  
 CC bones, to increase the limited storage viability of transplanted hearts,  
 CC kidneys, skin and other organs and tissues. The compounds are also useful  
 CC in the treatment of diseases of the central nervous system (including  
 CC AIDS dementia, stroke, amyotrophic lateral sclerosis), Parkinson's  
 CC disease, Huntington's disease, disease of the musculature, cardiac  
 CC fatigue of congestive heart failure, muscle weakness syndrome associated  
 CC with myopathies, neurological disorders, arthritis, systemic  
 CC hypertension, arteriosclerosis, oedema, septic shock, pulmonary  
 CC hypertension, impotence, infertility, endometriosis, premature uterine  
 CC contractions, microbial infections, gout, type II diabetes mellitus,  
 CC inflammation of the lungs, asthma, pneumonia, cystic fibrosis, chronic  
 CC sinusitis and autoimmune disease. The present sequence represents the  
 CC human EC-SOD protein. EC-SOD is a tetrameric glycosylated copper and zinc  
 CC containing enzyme. The enzyme is used to illustrate the activity of the  
 CC porphines of the invention.  
 XX  
 SQ Sequence 240 AA;

Query Match 1.8%; Score 7; DB 21; Length 240;  
 Best Local Similarity 100.0%; Pred. No. 2e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 358 PSATLDA 364  
 |||||  
 Db 67 PSATLDA 73

Search completed: December 25, 2002, 20:30:28  
 Job time : 46 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compen Ltd.

## OM protein - protein search, using sw model

Run on: December 25, 2002, 20:29:38 / Search time 17 Seconds  
(without alignments)  
692.304 Million cell updates/sec

Title: US-09-889-746-2

Perfect score: 400  
Sequence: 1 MHSHPQWLSDPLLSVAVT.....SKAEVLGRVINKPKRALF 400

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 262574 seque, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

## Database :

- Issued Patents AA:\*
- 1: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep:\*
  - 2: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep:\*
  - 3: /cgn2\_6/prodata/1/iaa/6A\_COMB.pep:\*
  - 4: /cgn2\_6/prodata/1/iaa/6B\_COMB.pep:\*
  - 5: /cgn2\_6/prodata/1/iaa/PCTUS\_COMB.pep:\*
  - 6: /cgn2\_6/prodata/1/iaa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	2.5	117	1 US-08-249-013-6	Sequence 6, Appl1
2	10	2.5	117	2 US-08-886-863-6	Sequence 6, Appl1
3	10	2.5	117	4 US-09-175-229-6	Sequence 6, Appl1
4	10	2.5	117	5 PCT-US95-06764-6	Sequence 6, Appl1
5	7	1.8	114	4 US-08-679-493A-187	Sequence 187, App
6	7	1.8	222	3 US-08-556-965-2	Sequence 2, Appl1
7	7	1.8	222	6 5472691-3	Patent No. 5472691
8	7	1.8	235	4 US-09-134-001C-4827	Sequence 4827, Ap
9	7	1.8	240	2 US-08-023-980B-45	Sequence 45, Appl
10	7	1.8	240	2 US-08-486-953A-53	Sequence 53, Appl
11	7	1.8	240	4 US-08-679-493A-186	Sequence 186, App
12	7	1.8	240	6 5472691-2	Patent No. 5472691
13	7	1.8	288	4 US-09-134-001C-3317	Sequence 3317, Ap
14	7	1.8	303	1 US-08-109-919A-2	Sequence 2, Appl1
15	7	1.8	303	1 US-08-459-019A-2	Sequence 2, Appl1
16	7	1.8	303	2 US-08-460-428A-2	Sequence 2, Appl1
17	7	1.8	303	3 US-08-458-860A-2	Sequence 2, Appl1
18	7	1.8	338	1 US-08-229-287-4	Sequence 4, Appl1
19	7	1.8	339	1 US-08-153-848-44	Sequence 4, Appl1
20	7	1.8	339	2 US-08-812-871-3	Sequence 3, Appl1
21	7	1.8	339	3 US-09-299-843A-44	Sequence 44, Appl
22	7	1.8	339	3 US-09-088-337B-44	Sequence 44, Appl
23	7	1.8	339	5 PCT-US93-11153-44	Sequence 44, Appl
24	7	1.8	339	5 PCT-US95-07180-2	Sequence 44, Appl
25	7	1.8	416	4 US-09-416-050A-4	Sequence 4, Appl1
26	7	1.8	416	4 US-09-664-800-4	Sequence 4, Appl1
27	7	1.8	416	4 US-09-665-309-4	Sequence 4, Appl1

28	7	1.8	416	4 US-09-661-569-4	Sequence 4, Appl1
29	7	1.8	426	4 US-08-676-444-40	Sequence 40, Appl
30	7	1.8	2763	3 US-08-496-944-2	Sequence 2, Appl1
31	6	1.5	11	4 US-09-041-886-43	Sequence 43, Appl
32	6	1.5	12	1 US-07-841-997A-42	Sequence 42, Appl
33	6	1.5	12	1 US-08-290-301-42	Sequence 42, Appl
34	6	1.5	12	4 US-09-013-598-42	Sequence 42, Appl
35	6	1.5	14	4 US-09-041-886-3	Sequence 4, Appl1
36	6	1.5	14	4 US-09-041-886-5	Sequence 4, Appl1
37	6	1.5	14	4 US-09-041-886-5	Sequence 4, Appl1
38	6	1.5	14	4 US-09-041-886-5	Sequence 4, Appl1
39	6	1.5	17	1 US-08-554-612C-45	Sequence 45, Appl
40	6	1.5	19	4 US-09-549-831-7	Sequence 7, Appl1
41	6	1.5	20	3 US-08-256-747C-9	Sequence 9, Appl1
42	6	1.5	20	4 US-08-834-130A-9	Sequence 9, Appl1
43	6	1.5	21	1 US-08-295-411-6	Sequence 6, Appl1
44	6	1.5	21	2 US-08-955-471-6	Sequence 6, Appl1
45	6	1.5	21	5 PCT-US92-10242-6	Sequence 6, Appl1
46	6	1.5	24	2 US-08-651-420A-22	Sequence 22, Appl
47	6	1.5	25	4 US-09-041-886-44	Sequence 44, Appl
48	6	1.5	27	1 US-08-378-761A-31	Sequence 31, Appl
49	6	1.5	27	1 US-08-485-286-31	Sequence 31, Appl
50	6	1.5	27	6 5248606-8	Patent No. 5248606
51	6	1.5	28	4 US-08-905-223-477	Sequence 477, App
52	6	1.5	28	4 US-09-041-886-37	Sequence 37, Appl
53	6	1.5	28	4 US-09-041-886-38	Sequence 38, Appl
54	6	1.5	28	4 US-09-041-886-39	Sequence 39, Appl
55	6	1.5	28	4 US-09-041-886-42	Sequence 42, Appl
56	6	1.5	28	4 US-09-041-886-53	Sequence 53, Appl
57	6	1.5	28	4 US-09-041-886-54	Sequence 54, Appl
58	6	1.5	31	1 US-08-324-301-11	Sequence 11, Appl
59	6	1.5	34	4 US-08-082-278B-381	Sequence 381, App
60	6	1.5	34	4 US-09-315-304B-381	Sequence 381, App
61	6	1.5	35	2 US-08-651-420A-23	Sequence 23, Appl
62	6	1.5	35	3 US-08-486-099-18	Sequence 18, Appl
63	6	1.5	35	3 US-08-486-099-33	Sequence 33, Appl
64	6	1.5	35	3 US-08-486-099-34	Sequence 34, Appl
65	6	1.5	35	3 US-08-486-099-35	Sequence 35, Appl
66	6	1.5	35	3 US-08-486-099-36	Sequence 36, Appl
67	6	1.5	35	3 US-08-486-099-37	Sequence 37, Appl
68	6	1.5	35	3 US-08-486-099-38	Sequence 38, Appl
69	6	1.5	35	3 US-08-486-099-39	Sequence 39, Appl
70	6	1.5	35	3 US-08-486-099-40	Sequence 40, Appl
71	6	1.5	35	3 US-08-486-099-41	Sequence 41, Appl
72	6	1.5	35	3 US-08-486-099-42	Sequence 42, Appl
73	6	1.5	35	3 US-08-486-099-43	Sequence 43, Appl
74	6	1.5	35	3 US-08-486-099-44	Sequence 44, Appl
75	6	1.5	35	3 US-08-486-099-45	Sequence 45, Appl
76	6	1.5	35	3 US-08-486-099-46	Sequence 46, Appl
77	6	1.5	35	3 US-08-486-099-47	Sequence 47, Appl
78	6	1.5	35	3 US-08-486-099-48	Sequence 48, Appl
79	6	1.5	35	3 US-08-486-099-49	Sequence 49, Appl
80	6	1.5	35	3 US-08-360-107A-18	Sequence 18, Appl
81	6	1.5	35	3 US-08-360-107A-38	Sequence 38, Appl
82	6	1.5	35	3 US-08-360-107A-34	Sequence 34, Appl
83	6	1.5	35	3 US-08-360-107A-35	Sequence 35, Appl
84	6	1.5	35	3 US-08-360-107A-36	Sequence 36, Appl
85	6	1.5	35	3 US-08-360-107A-37	Sequence 37, Appl
86	6	1.5	35	3 US-08-360-107A-38	Sequence 38, Appl
87	6	1.5	35	3 US-08-360-107A-39	Sequence 39, Appl
88	6	1.5	35	3 US-08-360-107A-40	Sequence 40, Appl
89	6	1.5	35	3 US-08-360-107A-41	Sequence 41, Appl
90	6	1.5	35	3 US-08-360-107A-42	Sequence 42, Appl
91	6	1.5	35	3 US-08-360-107A-43	Sequence 43, Appl
92	6	1.5	35	3 US-08-360-107A-44	Sequence 44, Appl
93	6	1.5	35	3 US-08-360-107A-45	Sequence 45, Appl
94	6	1.5	35	3 US-08-360-107A-46	Sequence 46, Appl
95	6	1.5	35	3 US-08-360-107A-47	Sequence 47, Appl
96	6	1.5	35	3 US-08-360-107A-48	Sequence 48, Appl
97	6	1.5	35	3 US-08-484-223B-18	Sequence 18, Appl
98	6	1.5	35	3 US-08-484-223B-33	Sequence 33, Appl
99	6	1.5	35	3 US-08-484-223B-34	Sequence 34, Appl
100	6	1.5	35	3 US-08-484-223B-35	Sequence 35, Appl



247	6	1.5	35	4	US-08-474-349A-39	Sequence 39, App1	320	6	1.5	44	5	PCT-US92-03089-2	Sequence 2, App1
248	6	1.5	35	4	US-08-474-349A-40	Sequence 40, App1	321	6	1.5	45	1	US-07-749-541A-4	Sequence 4, App1
249	6	1.5	35	4	US-08-474-349A-41	Sequence 41, App1	322	6	1.5	45	1	US-08-275-327-1	Sequence 1, App1
250	6	1.5	35	4	US-08-474-349A-42	Sequence 42, App1	323	6	1.5	45	5	PCT-US92-03089-1	Sequence 1, App1
251	6	1.5	35	4	US-08-474-349A-43	Sequence 43, App1	324	6	1.5	56	3	US-08-486-099-101	Sequence 10, App
252	6	1.5	35	4	US-08-474-349A-44	Sequence 44, App1	325	6	1.5	56	3	US-08-360-107A-90	Sequence 90, App1
253	6	1.5	35	4	US-08-474-349A-45	Sequence 45, App1	326	6	1.5	56	3	US-08-360-110A-111	Sequence 111, App
254	6	1.5	35	4	US-08-474-349A-46	Sequence 46, App1	327	6	1.5	56	3	US-08-484-222B-101	Sequence 101, App
255	6	1.5	35	4	US-08-474-349A-47	Sequence 47, App1	328	6	1.5	56	3	US-08-919-597-101	Sequence 101, App
256	6	1.5	35	4	US-08-474-349A-48	Sequence 48, App1	329	6	1.5	56	3	US-08-475-668A-101	Sequence 101, App
257	6	1.5	35	4	US-08-474-349A-153	Sequence 153, App	330	6	1.5	56	3	US-08-485-551A-101	Sequence 101, App
258	6	1.5	35	4	US-08-474-349A-321	Sequence 321, App	331	6	1.5	56	3	US-08-471-913A-101	Sequence 101, App
259	6	1.5	35	4	US-08-474-349A-322	Sequence 322, App	332	6	1.5	56	4	US-08-485-264A-101	Sequence 101, App
260	6	1.5	35	4	US-08-474-349A-323	Sequence 323, App	333	6	1.5	56	4	US-08-474-349A-101	Sequence 101, App
261	6	1.5	35	4	US-08-474-349A-324	Sequence 324, App	334	6	1.5	56	4	US-08-894-626-94	Sequence 4, App1
262	6	1.5	35	4	US-08-474-349A-325	Sequence 325, App	335	6	1.5	56	4	US-08-255-208A-35	Sequence 35, App1
263	6	1.5	35	4	US-08-474-349A-326	Sequence 326, App	336	6	1.5	57	1	US-08-241-853-3	Sequence 3, App1
264	6	1.5	35	4	US-08-474-349A-327	Sequence 327, App	337	6	1.5	57	2	US-08-850-917-3	Sequence 3, App1
265	6	1.5	35	4	US-08-474-349A-328	Sequence 328, App	338	6	1.5	68	3	US-08-995-050-1	Sequence 1, App1
266	6	1.5	35	4	US-08-474-349A-329	Sequence 329, App	339	6	1.5	79	2	US-09-196-048-2	Sequence 2, App1
267	6	1.5	35	4	US-08-474-349A-330	Sequence 330, App	340	6	1.5	79	4	US-09-127-946-10	Sequence 6, App1
268	6	1.5	35	4	US-08-474-349A-331	Sequence 331, App	341	6	1.5	84	4	US-09-404-671-6	Sequence 6, App1
269	6	1.5	35	4	US-08-474-349A-332	Sequence 332, App	342	6	1.5	86	4	US-09-134-001C-3816	Sequence 3816, App
270	6	1.5	35	4	US-08-474-349A-333	Sequence 333, App	343	6	1.5	92	4	US-09-127-946-8	Sequence 8, App1
271	6	1.5	35	4	US-08-474-349A-334	Sequence 334, App	344	6	1.5	93	4	US-09-247-155-114	Sequence 114, App
272	6	1.5	35	4	US-08-474-349A-335	Sequence 335, App	345	6	1.5	96	1	US-08-565-386-23	Sequence 23, App1
273	6	1.5	35	4	US-08-474-349A-336	Sequence 336, App	346	6	1.5	98	2	US-09-010-928B-25	Sequence 25, App1
274	6	1.5	35	4	US-08-474-349A-337	Sequence 337, App	347	6	1.5	101	4	US-09-248-528-13	Sequence 13, App1
275	6	1.5	35	4	US-08-474-349A-338	Sequence 338, App	348	6	1.5	101	4	US-09-549-108-13	Sequence 13, App1
276	6	1.5	35	4	US-08-474-349A-339	Sequence 339, App	349	6	1.5	101	4	US-09-549-111-13	Sequence 13, App1
277	6	1.5	35	4	US-08-474-349A-340	Sequence 340, App	350	6	1.5	101	4	US-09-549-106-13	Sequence 13, App1
278	6	1.5	35	4	US-08-474-349A-341	Sequence 341, App	351	6	1.5	101	4	US-09-550-394-13	Sequence 13, App1
279	6	1.5	35	4	US-08-474-349A-342	Sequence 342, App	352	6	1.5	105	1	US-07-998-003A-99	Sequence 99, App1
280	6	1.5	35	4	US-09-315-304B-39	Sequence 39, App1	353	6	1.5	105	1	US-08-453-274B-99	Sequence 99, App1
281	6	1.5	35	4	US-09-315-304B-79	Sequence 79, App1	354	6	1.5	105	1	US-08-326-117B-4	Sequence 4, App1
282	6	1.5	35	4	US-09-315-304B-80	Sequence 80, App1	355	6	1.5	105	1	US-08-453-695A-99	Sequence 99, App1
283	6	1.5	35	4	US-09-315-304B-186	Sequence 186, App1	356	6	1.5	105	1	US-08-268-161A-99	Sequence 99, App1
284	6	1.5	35	4	US-09-315-304B-187	Sequence 187, App1	357	6	1.5	105	2	US-08-453-702A-99	Sequence 99, App1
285	6	1.5	35	4	US-09-315-304B-188	Sequence 188, App1	358	6	1.5	105	3	US-08-983-129-4	Sequence 4, App1
286	6	1.5	35	4	US-09-315-304B-189	Sequence 189, App1	359	6	1.5	105	4	US-09-099-639-99	Sequence 99, App1
287	6	1.5	35	4	US-09-315-304B-190	Sequence 190, App1	360	6	1.5	105	4	US-09-178-176B-9	Sequence 9, App1
288	6	1.5	35	4	US-09-315-304B-191	Sequence 191, App1	361	6	1.5	105	4	US-09-457-864-9	Sequence 9, App1
289	6	1.5	35	4	US-09-315-304B-192	Sequence 192, App1	362	6	1.5	105	5	PCT-US93-12588-99	Sequence 99, App1
290	6	1.5	35	4	US-09-315-304B-193	Sequence 193, App1	363	6	1.5	105	5	PCT-US95-08071-99	Sequence 99, App1
291	6	1.5	35	4	US-09-315-304B-194	Sequence 194, App1	364	6	1.5	117	4	US-08-936-165A-86	Sequence 86, App
292	6	1.5	35	4	US-09-315-304B-195	Sequence 195, App1	365	6	1.5	117	4	US-09-249-542-12	Sequence 12, App1
293	6	1.5	35	4	US-09-315-304B-196	Sequence 196, App1	366	6	1.5	150	4	US-09-134-001C-3023	Sequence 3023, App
294	6	1.5	35	4	US-09-315-304B-197	Sequence 197, App1	367	6	1.5	163	4	US-09-354-129-12	Sequence 12, App1
295	6	1.5	35	4	US-09-315-304B-198	Sequence 198, App1	368	6	1.5	163	4	US-09-504-357-12	Sequence 12, App1
296	6	1.5	35	4	US-09-315-304B-199	Sequence 199, App1	369	6	1.5	168	1	US-08-441-139-10	Sequence 10, App1
297	6	1.5	35	4	US-09-315-304B-200	Sequence 200, App1	370	6	1.5	170	4	US-09-523-433B-24	Sequence 24, App1
298	6	1.5	35	4	US-09-315-304B-201	Sequence 201, App1	371	6	1.5	186	4	US-09-134-001C-4741	Sequence 4741, App
299	6	1.5	35	4	US-09-315-304B-202	Sequence 202, App1	372	6	1.5	189	2	US-08-837-029-2	Sequence 2, App1
300	6	1.5	35	4	US-09-315-304B-203	Sequence 203, App1	373	6	1.5	191	4	US-09-183-861-82	Sequence 82, App1
301	6	1.5	35	4	US-09-315-304B-204	Sequence 204, App1	374	6	1.5	191	4	US-09-022-765-82	Sequence 82, App1
302	6	1.5	35	4	US-09-315-304B-205	Sequence 205, App1	375	6	1.5	198	4	US-09-325-932A-151	Sequence 151, App
303	6	1.5	35	4	US-09-315-304B-206	Sequence 206, App1	376	6	1.5	199	1	US-08-602-262-2	Sequence 2, App1
304	6	1.5	35	4	US-09-315-304B-207	Sequence 207, App1	377	6	1.5	199	3	US-09-004-716-2	Sequence 2, App1
305	6	1.5	35	4	US-09-315-304B-548	Sequence 548, App1	378	6	1.5	199	4	US-08-894-251A-2	Sequence 2, App1
306	6	1.5	35	4	US-09-315-304B-863	Sequence 863, App1	379	6	1.5	204	2	US-08-531-525-32	Sequence 32, App1
307	6	1.5	35	4	US-09-315-304B-864	Sequence 864, App1	380	6	1.5	204	2	US-08-718-270A-32	Sequence 32, App1
308	6	1.5	35	4	US-09-315-304B-880	Sequence 880, App1	381	6	1.5	204	4	US-09-549-831-6	Sequence 6, App1
309	6	1.5	35	4	US-09-315-304B-881	Sequence 881, App1	382	6	1.5	212	2	US-08-761-248B-2	Sequence 2, App1
310	6	1.5	35	4	US-09-315-304B-882	Sequence 882, App1	383	6	1.5	212	2	US-08-761-248B-6	Sequence 6, App1
311	6	1.5	35	4	US-09-315-304B-883	Sequence 883, App1	384	6	1.5	214	4	US-08-861-774B-32	Sequence 32, App1
312	6	1.5	35	4	US-09-315-304B-888	Sequence 888, App1	385	6	1.5	221	2	US-08-943-087-52	Sequence 52, App1
313	6	1.5	35	4	US-09-315-304B-889	Sequence 889, App1	386	6	1.5	221	2	US-08-943-087-54	Sequence 54, App1
314	6	1.5	37	4	US-09-315-304B-916	Sequence 916, App1	387	6	1.5	221	2	US-08-943-087-55	Sequence 55, App1
315	6	1.5	41	2	US-08-255-208A-18	Sequence 18, App1	388	6	1.5	221	2	US-08-943-087-56	Sequence 56, App1
316	6	1.5	44	1	US-08-726-306A-83	Sequence 83, App1	389	6	1.5	221	2	US-08-943-087-58	Sequence 58, App1
317	6	1.5	44	1	US-07-749-541A-7	Sequence 7, App1	390	6	1.5	221	2	US-08-943-087-60	Sequence 60, App1
318	6	1.5	44	1	US-08-275-327-2	Sequence 2, App1	391	6	1.5	228	1	US-08-278-091-10	Sequence 10, App1
319	6	1.5	44	4	US-09-149-476-654	Sequence 654, App1	392	6	1.5	228	1	US-08-483-859-10	Sequence 10, App1

333	6	1.5	228	1	US-08-472-173-10	Sequence 10, Appl	466	6	1.5	338	4	US-09-422-662-4	Sequence 4, Appl
334	6	1.5	228	2	US-08-487-167-10	Sequence 10, Appl	467	6	1.5	338	4	US-09-422-662-37	Sequence 37, Appl
335	6	1.5	228	2	US-08-482-816-10	Sequence 10, Appl	468	6	1.5	340	4	US-09-651-000-2	Sequence 2, Appl
336	6	1.5	228	2	US-08-296-149-10	Sequence 10, Appl	469	6	1.5	348	4	US-09-134-001C-4857	Sequence 4857, Ap
337	6	1.5	228	2	US-08-801-499-10	Sequence 10, Appl	470	6	1.5	351	2	US-08-701-191A-3	Sequence 3, Appl
338	6	1.5	228	2	US-08-615-271-10	Sequence 10, Appl	471	6	1.5	351	4	US-09-916-109-7	Sequence 7, Appl
339	6	1.5	228	3	US-09-074-660-10	Sequence 10, Appl	472	6	1.5	356	4	US-09-071-035-378	Sequence 378, App
340	6	1.5	228	3	US-09-074-659-10	Sequence 10, Appl	473	6	1.5	359	2	US-08-560-398-10	Sequence 10, Appl
341	6	1.5	228	3	US-09-106-468-10	Sequence 10, Appl	474	6	1.5	359	4	US-09-134-001C-4760	Sequence 4760, Ap
342	6	1.5	228	4	US-09-106-466A-10	Sequence 10, Appl	475	6	1.5	366	1	US-08-700-359-22	Sequence 22, Appl
343	6	1.5	228	4	US-09-106-467-10	Sequence 10, Appl	476	6	1.5	371	2	US-08-378-617A-12	Sequence 12, Appl
344	6	1.5	235	4	US-09-049-672A-12	Sequence 12, Appl	477	6	1.5	372	2	US-08-360-606B-31	Sequence 31, Appl
345	6	1.5	236	1	US-08-315-695-22	Sequence 22, Appl	478	6	1.5	375	4	US-09-134-001C-5050	Sequence 5050, Ap
346	6	1.5	245	2	US-08-874-138-2	Sequence 2, Appl	479	6	1.5	375	4	US-09-323-872A-33	Sequence 33, Appl
347	6	1.5	245	4	US-08-879-941-4	Sequence 4, Appl	480	6	1.5	380	4	US-09-877-730-4	Sequence 4, Appl
348	6	1.5	245	4	US-09-747-116-4	Sequence 4, Appl	481	6	1.5	387	4	US-09-916-109-6	Sequence 6, Appl
349	6	1.5	245	4	US-09-325-932A-89	Sequence 89, Appl	482	6	1.5	392	1	US-08-723-938-1	Sequence 1, Appl
350	6	1.5	247	1	US-08-488-113B-6	Sequence 6, Appl	483	6	1.5	392	2	US-09-080-538-1	Sequence 1, Appl
351	6	1.5	247	1	US-08-477-484B-6	Sequence 6, Appl	484	6	1.5	394	1	US-07-914-281-4	Sequence 4, Appl
352	6	1.5	247	2	US-08-646-360-6	Sequence 6, Appl	485	6	1.5	394	1	US-08-353-246-4	Sequence 4, Appl
353	6	1.5	247	4	US-08-833-765-6	Sequence 6, Appl	486	6	1.5	394	1	US-08-525-058A-4	Sequence 4, Appl
354	6	1.5	247	4	US-09-136-389-6	Sequence 6, Appl	487	6	1.5	394	2	US-08-696-731-4	Sequence 4, Appl
355	6	1.5	247	4	US-09-610-838-6	Sequence 6, Appl	488	6	1.5	394	4	US-09-042-531-4	Sequence 4, Appl
356	6	1.5	252	4	US-09-134-001C-5644	Sequence 5644, Ap	489	6	1.5	394	4	US-08-379-040-2	Sequence 2, Appl
357	6	1.5	255	1	US-07-901-707-6	Sequence 6, Appl	490	6	1.5	394	5	PCT-US91-00899-9	Sequence 9, Appl
358	6	1.5	255	1	US-07-988-430-6	Sequence 6, Appl	491	6	1.5	398	3	US-08-931-220-5	Sequence 5, Appl
359	6	1.5	255	1	US-08-425-336-6	Sequence 6, Appl	492	6	1.5	398	5	PCT-US95-11723-5	Sequence 5, Appl
360	6	1.5	255	5	PCT-US92-09487-6	Sequence 6, Appl	493	6	1.5	398	5	PCT-US96-05997-1	Sequence 1, Appl
361	6	1.5	267	1	US-08-378-761A-74	Sequence 74, Appl	494	6	1.5	411	2	US-08-773-870-1	Sequence 1, Appl
362	6	1.5	267	1	US-08-485-286-74	Sequence 74, Appl	495	6	1.5	416	3	US-08-910-505-2	Sequence 2, Appl
363	6	1.5	269	1	US-07-706-691G-8	Sequence 8, Appl	496	6	1.5	416	3	US-08-910-505-4	Sequence 4, Appl
364	6	1.5	269	1	US-08-254-021-8	Sequence 8, Appl	497	6	1.5	416	4	US-09-493-459-2	Sequence 2, Appl
365	6	1.5	269	2	US-08-618-446-8	Sequence 8, Appl	498	6	1.5	416	4	US-09-493-459-4	Sequence 4, Appl
366	6	1.5	269	4	US-09-580-135-8	Sequence 8, Appl	499	6	1.5	427	4	US-09-086-483A-4	Sequence 4, Appl
367	6	1.5	269	4	US-09-585-798-8	Sequence 8, Appl	500	6	1.5	427	4	US-09-041-886-2	Sequence 2, Appl
368	6	1.5	277	2	US-08-469-537A-57	Sequence 57, Appl	501	6	1.5	427	4	US-09-006-353A-5	Sequence 5, Appl
369	6	1.5	283	1	US-08-583-672-2	Sequence 2, Appl	502	6	1.5	427	4	US-08-486-099-95	Sequence 95, Appl
370	6	1.5	283	2	US-08-202-044-2	Sequence 2, Appl	503	6	1.5	430	3	US-08-360-107A-105	Sequence 105, App
371	6	1.5	283	4	US-08-751-344B-2	Sequence 2, Appl	504	6	1.5	430	3	US-08-484-223B-95	Sequence 95, Appl
372	6	1.5	287	1	US-08-624-125-17	Sequence 17, Appl	505	6	1.5	430	3	US-08-919-597-95	Sequence 95, Appl
373	6	1.5	289	1	US-07-923-692C-4	Sequence 4, Appl	506	6	1.5	430	3	US-08-475-668A-95	Sequence 95, Appl
374	6	1.5	289	1	US-08-184-237-4	Sequence 4, Appl	507	6	1.5	430	3	US-08-485-551A-95	Sequence 95, Appl
375	6	1.5	289	2	US-08-482-920-4	Sequence 4, Appl	508	6	1.5	430	3	US-08-471-513A-95	Sequence 95, Appl
376	6	1.5	289	3	US-08-484-321-4	Sequence 4, Appl	509	6	1.5	430	3	US-08-485-264A-95	Sequence 95, Appl
377	6	1.5	289	4	US-08-483-502-4	Sequence 4, Appl	510	6	1.5	430	4	US-08-474-349A-95	Sequence 95, Appl
378	6	1.5	289	4	US-09-726-651A-4	Sequence 40, Appl	511	6	1.5	430	4	US-08-255-208A-31	Sequence 31, Appl
379	6	1.5	294	4	US-09-372-422A-40	Sequence 38, Appl	512	6	1.5	434	2	US-09-106-075A-86	Sequence 86, Appl
380	6	1.5	295	4	US-09-134-001C-5221	Sequence 5221, Ap	513	6	1.5	434	2	US-08-677-049-7	Sequence 7, Appl
381	6	1.5	302	4	US-09-877-730-14	Sequence 14, Appl	514	6	1.5	434	2	US-08-710-249-4	Sequence 4, Appl
382	6	1.5	311	4	US-09-438-833-7	Sequence 7, Appl	515	6	1.5	434	4	US-09-220-157A-4	Sequence 4, Appl
383	6	1.5	312	4	US-09-821-736-5	Sequence 5, Appl	516	6	1.5	441	4	US-09-651-200-4	Sequence 4, Appl
384	6	1.5	314	4	US-09-071-035-380	Sequence 380, App	517	6	1.5	441	4	US-09-916-109-5	Sequence 5, Appl
385	6	1.5	315	2	US-08-701-191A-2	Sequence 2, Appl	518	6	1.5	447	4	US-09-134-001C-5311	Sequence 5311, Ap
386	6	1.5	315	4	US-09-134-001C-4141	Sequence 4, Appl	519	6	1.5	461	4	US-09-134-001C-3235	Sequence 3235, Ap
387	6	1.5	316	1	US-08-403-634-4	Sequence 4, Appl	520	6	1.5	462	2	US-08-865-597A-2	Sequence 2, Appl
388	6	1.5	316	1	US-08-403-634-31	Sequence 31, Appl	521	6	1.5	462	4	US-09-189-527-13	Sequence 13, Appl
389	6	1.5	316	4	US-08-913-441B-4	Sequence 4, Appl	522	6	1.5	465	4	US-09-292-097-15	Sequence 15, Appl
390	6	1.5	316	4	US-08-913-441B-31	Sequence 31, Appl	523	6	1.5	467	1	US-09-134-001C-3235	Sequence 33, Appl
391	6	1.5	317	1	US-07-688-299-1	Sequence 1, Appl	524	6	1.5	476	1	US-08-216-276A-33	Sequence 33, Appl
392	6	1.5	317	1	US-07-688-299-1	Sequence 1, Appl	525	6	1.5	481	1	US-07-927-071-1	Sequence 1, Appl
393	6	1.5	317	1	US-07-980-517A-1	Sequence 3, Appl	526	6	1.5	481	6	5164481-1	Patent No. 5164481
394	6	1.5	318	1	US-07-688-299-3	Sequence 125, Appl	527	6	1.5	483	4	US-09-916-109-4	Sequence 4, Appl
395	6	1.5	318	1	US-07-688-299-3	Sequence 13, Appl	528	6	1.5	484	6	517673-8	Patent No. 517673
396	6	1.5	318	4	US-09-060-756-727	Sequence 727, App	529	6	1.5	485	1	US-07-881-075-1	Sequence 1, Appl
397	6	1.5	318	4	US-09-601-027-1	Sequence 1, Appl	530	6	1.5	485	1	US-08-120-827-1	Sequence 1, Appl
398	6	1.5	320	1	US-08-362-739-2	Sequence 2, Appl	531	6	1.5	485	1	US-08-478-675-1	Sequence 1, Appl
399	6	1.5	321	4	US-09-171-461-22	Sequence 22, Appl	532	6	1.5	487	6	517673-6	Patent No. 517673
400	6	1.5	330	4	US-09-188-930-125	Sequence 125, App	533	6	1.5	499	4	US-09-457-040B-13	Sequence 13, Appl
401	6	1.5	330	4	US-09-615-192A-284	Sequence 284, App	534	6	1.5	504	1	US-08-135-511-30	Sequence 30, Appl
402	6	1.5	333	4	US-09-171-710-2	Sequence 2, Appl	535	6	1.5	504	1	US-08-187-453-30	Sequence 30, Appl
403	6	1.5	334	4	US-09-060-756-728	Sequence 728, App	536	6	1.5	505	4	US-09-729-995-2	Sequence 2, Appl
404	6	1.5	338	3	US-08-917-299-4	Sequence 4, Appl	537	6	1.5	505	4	US-09-729-995-4	Sequence 4, Appl
405	6	1.5	338	3	US-08-917-299-37	Sequence 37, Appl	538	6	1.5	506	4	US-08-942-012B-25	Sequence 25, Appl

539	6	1.5	506	6	5180581-2	Patent No. 5180581	612	6	1.5	607	2	US-08-472-534-5	Sequence 5, Appl1
540	6	1.5	512	4	US-09-390-234-24	Sequence 24, Appl1	613	6	1.5	608	4	US-09-413-814-92	Sequence 92, Appl1
541	6	1.5	519	4	US-09-008-271A-9	Sequence 9, Appl1	614	6	1.5	613	4	US-09-438-833-6	Sequence 6, Appl1
542	6	1.5	520	4	US-09-149-476-351	Sequence 351, App	615	6	1.5	615	4	US-09-462-844-3	Sequence 3, Appl1
543	6	1.5	521	4	US-09-134-001C-5639	Sequence 5639, Ap	616	6	1.5	617	4	US-09-188-930-303	Sequence 303, App
544	6	1.5	523	3	US-08-651-999A-2	Sequence 2, Appl1	617	6	1.5	652	4	US-09-438-833-5	Sequence 5, Appl1
545	6	1.5	525	4	US-09-385-752-2	Sequence 2, Appl1	618	6	1.5	654	1	US-08-487-890A-12	Sequence 12, Appl1
546	6	1.5	525	1	US-08-077-939-19	Sequence 19, Appl1	619	6	1.5	654	2	US-08-478-435-12	Sequence 12, Appl1
547	6	1.5	525	1	US-08-461-599-19	Sequence 19, Appl1	620	6	1.5	654	2	US-08-337-483-12	Sequence 12, Appl1
548	6	1.5	525	1	US-08-461-621-19	Sequence 19, Appl1	621	6	1.5	654	2	US-08-478-373-12	Sequence 12, Appl1
549	6	1.5	525	1	US-08-465-334-19	Sequence 19, Appl1	622	6	1.5	654	3	US-08-474-671-12	Sequence 12, Appl1
550	6	1.5	525	2	US-08-811-897A-23	Sequence 23, Appl1	623	6	1.5	654	3	US-08-483-577A-12	Sequence 12, Appl1
551	6	1.5	525	2	US-08-853-213-23	Sequence 23, Appl1	624	6	1.5	654	4	US-08-897-438-12	Sequence 12, Appl1
552	6	1.5	525	4	US-09-201-474-23	Sequence 23, Appl1	625	6	1.5	654	4	US-08-637-654-12	Sequence 12, Appl1
553	6	1.5	529	2	US-08-836-442-3	Sequence 3, Appl1	626	6	1.5	654	4	US-08-649-518-12	Sequence 12, Appl1
554	6	1.5	534	4	US-09-651-200-6	Sequence 6, Appl1	627	6	1.5	697	2	US-08-674-351-4	Sequence 4, Appl1
555	6	1.5	534	4	US-09-651-200-24	Sequence 24, Appl1	628	6	1.5	701	4	US-08-922-511-2	Sequence 2, Appl1
556	6	1.5	539	2	US-08-467-963C-2	Sequence 2, Appl1	629	6	1.5	701	4	US-09-416-874A-2	Sequence 2, Appl1
557	6	1.5	539	2	US-08-838-189D-2	Sequence 2, Appl1	630	6	1.5	716	4	US-08-971-188-11	Sequence 5, Appl1
558	6	1.5	539	3	US-08-852-344D-2	Sequence 2, Appl1	631	6	1.5	716	4	US-08-971-188-11	Sequence 11, Appl1
559	6	1.5	539	3	US-08-344-639B-2	Sequence 2, Appl1	632	6	1.5	729	4	US-09-625-188-20	Sequence 20, Appl1
560	6	1.5	539	4	US-08-467-969A-2	Sequence 2, Appl1	633	6	1.5	737	4	US-08-971-188-4	Sequence 4, Appl1
561	6	1.5	539	4	US-08-467-961A-2	Sequence 2, Appl1	634	6	1.5	745	2	US-09-010-928B-28	Sequence 28, Appl1
562	6	1.5	539	4	US-08-001-584A-2	Sequence 2, Appl1	635	6	1.5	750	3	US-08-814-052-2	Sequence 2, Appl1
563	6	1.5	542	1	US-08-321-587-2	Sequence 2, Appl1	636	6	1.5	750	3	US-08-812-829-2	Sequence 2, Appl1
564	6	1.5	549	4	US-09-291-922-30	Sequence 30, Appl1	637	6	1.5	756	4	US-09-438-833-11	Sequence 11, Appl1
565	6	1.5	550	4	US-09-330-245A-2	Sequence 2, Appl1	638	6	1.5	763	2	US-08-677-862-2	Sequence 2, Appl1
566	6	1.5	552	2	US-08-811-897A-27	Sequence 27, Appl1	639	6	1.5	763	2	US-09-252-571-2	Sequence 2, Appl1
567	6	1.5	552	2	US-08-855-213-27	Sequence 27, Appl1	640	6	1.5	763	3	US-09-434-065-2	Sequence 4, Appl1
568	6	1.5	552	4	US-09-201-474-27	Sequence 27, Appl1	641	6	1.5	763	3	US-08-789-275-4	Sequence 4, Appl1
569	6	1.5	553	2	US-08-811-897A-25	Sequence 25, Appl1	642	6	1.5	763	4	US-08-789-275-5	Sequence 5, Appl1
570	6	1.5	553	2	US-08-811-897A-29	Sequence 29, Appl1	643	6	1.5	764	1	US-08-375-300-4	Sequence 4, Appl1
571	6	1.5	553	2	US-08-855-213-25	Sequence 25, Appl1	644	6	1.5	764	3	US-09-177-431-4	Sequence 4, Appl1
572	6	1.5	553	2	US-08-855-213-29	Sequence 29, Appl1	645	6	1.5	764	5	PCT-US95-16930-4	Sequence 4, Appl1
573	6	1.5	553	2	US-08-943-087-2	Sequence 2, Appl1	646	6	1.5	794	4	US-09-134-001C-310	Sequence 4310, Ap
574	6	1.5	553	2	US-08-943-087-14	Sequence 14, Appl1	647	6	1.5	796	4	US-08-868-699A-2	Sequence 2, Appl1
575	6	1.5	553	2	US-08-943-087-16	Sequence 16, Appl1	648	6	1.5	796	4	US-09-757-014-2	Sequence 2, Appl1
576	6	1.5	553	2	US-08-943-087-18	Sequence 18, Appl1	649	6	1.5	805	2	US-08-480-473B-4	Sequence 4, Appl1
577	6	1.5	553	2	US-08-943-087-20	Sequence 20, Appl1	650	6	1.5	805	3	US-08-915-213-4	Sequence 4, Appl1
578	6	1.5	553	2	US-08-943-087-22	Sequence 22, Appl1	651	6	1.5	805	4	US-09-235-217-4	Sequence 4, Appl1
579	6	1.5	553	2	US-08-943-087-24	Sequence 24, Appl1	652	6	1.5	805	5	PCT-US96-10251-4	Sequence 4, Appl1
580	6	1.5	553	2	US-08-943-087-26	Sequence 26, Appl1	653	6	1.5	813	4	US-09-438-833-12	Sequence 12, Appl1
581	6	1.5	553	2	US-08-943-087-28	Sequence 28, Appl1	654	6	1.5	826	1	US-08-785-241-6	Sequence 6, Appl1
582	6	1.5	553	2	US-08-943-087-30	Sequence 30, Appl1	655	6	1.5	826	2	US-08-480-473B-2	Sequence 2, Appl1
583	6	1.5	553	2	US-08-943-087-32	Sequence 32, Appl1	656	6	1.5	826	3	US-08-915-213-3	Sequence 3, Appl1
584	6	1.5	553	2	US-08-943-087-34	Sequence 34, Appl1	657	6	1.5	826	3	US-09-148-547-2	Sequence 2, Appl1
585	6	1.5	553	2	US-08-943-087-36	Sequence 36, Appl1	658	6	1.5	826	4	US-09-235-217-2	Sequence 2, Appl1
586	6	1.5	553	2	US-08-943-087-38	Sequence 38, Appl1	659	6	1.5	826	4	US-09-380-662-23	Sequence 23, Appl1
587	6	1.5	553	2	US-08-943-087-40	Sequence 40, Appl1	660	6	1.5	826	4	US-09-438-833-1	Sequence 1, Appl1
588	6	1.5	553	2	US-08-943-087-42	Sequence 42, Appl1	661	6	1.5	826	4	US-09-877-730-16	Sequence 16, Appl1
589	6	1.5	553	2	US-08-943-087-44	Sequence 44, Appl1	662	6	1.5	826	5	PCT-US96-10251-2	Sequence 2, Appl1
590	6	1.5	553	2	US-08-943-087-46	Sequence 46, Appl1	663	6	1.5	834	1	US-07-977-434-8	Sequence 8, Appl1
591	6	1.5	553	2	US-08-943-087-48	Sequence 48, Appl1	664	6	1.5	834	1	US-07-977-434-10	Sequence 10, Appl1
592	6	1.5	553	4	US-09-201-474-25	Sequence 25, Appl1	665	6	1.5	834	1	US-08-077-384C-6	Sequence 6, Appl1
593	6	1.5	553	4	US-09-201-474-25	Sequence 25, Appl1	666	6	1.5	834	1	US-08-254-359A-6	Sequence 6, Appl1
594	6	1.5	560	5	PCT-US95-14418-5	Sequence 5, Appl1	667	6	1.5	834	1	US-08-384-490-31	Sequence 31, Appl1
595	6	1.5	560	5	PCT-US95-15337-5	Sequence 5, Appl1	668	6	1.5	834	1	US-08-483-043-6	Sequence 6, Appl1
596	6	1.5	574	4	US-09-134-001C-3382	Sequence 3382, Ap	669	6	1.5	834	1	US-08-491-337-2	Sequence 2, Appl1
597	6	1.5	575	1	US-07-683-957B-2	Sequence 3, Appl1	670	6	1.5	834	1	US-08-459-383-31	Sequence 31, Appl1
598	6	1.5	579	4	US-09-171-710-4	Sequence 4, Appl1	671	6	1.5	834	1	US-08-458-819-8	Sequence 8, Appl1
599	6	1.5	586	4	US-09-040-725A-1	Sequence 6, Appl1	672	6	1.5	834	1	US-08-458-819-10	Sequence 10, Appl1
600	6	1.5	586	4	US-08-591-079-8	Sequence 8, Appl1	673	6	1.5	834	1	US-08-481-238-6	Sequence 6, Appl1
601	6	1.5	593	2	US-08-591-079-8	Sequence 8, Appl1	674	6	1.5	834	2	US-08-471-066B-6	Sequence 6, Appl1
602	6	1.5	593	2	US-08-591-079-10	Sequence 10, Appl1	675	6	1.5	834	2	US-08-484-956-6	Sequence 6, Appl1
603	6	1.5	597	5	PCT-US95-14418-4	Sequence 4, Appl1	676	6	1.5	834	2	US-08-757-653-6	Sequence 6, Appl1
604	6	1.5	597	5	PCT-US95-15327-4	Sequence 4, Appl1	677	6	1.5	834	2	US-08-599-491-6	Sequence 6, Appl1
605	6	1.5	599	4	US-09-377-155-15	Sequence 15, Appl1	678	6	1.5	834	2	US-08-756-386-6	Sequence 6, Appl1
606	6	1.5	599	4	US-09-669-974-15	Sequence 15, Appl1	679	6	1.5	834	2	US-08-823-516-6	Sequence 6, Appl1
607	6	1.5	602	1	US-08-295-814B-2	Sequence 2, Appl1	680	6	1.5	834	3	US-08-683-853A-6	Sequence 6, Appl1
608	6	1.5	602	1	US-08-291-299-9	Sequence 2, Appl1	681	6	1.5	834	3	US-08-758-038-6	Sequence 6, Appl1
609	6	1.5	602	4	US-09-343-361-2	Sequence 2, Appl1	682	6	1.5	834	3	US-08-758-314-6	Sequence 6, Appl1
610	6	1.5	602	5	PCT-US93-01959-2	Sequence 2, Appl1	683	6	1.5	834	3	US-08-968-633-2	Sequence 2, Appl1
611	6	1.5	602	5	PCT-US95-10579-9	Sequence 9, Appl1	684	6	1.5	834	3	US-09-196-466-2	Sequence 2, Appl1

685	6	1.5	834	4	US-09-350-309-6	Sequence 6, Appli	758	6	1.5	1070	3	US-08-613-009A-11	Sequence 11, Appli
686	6	1.5	834	4	US-08-520-946-6	Sequence 6, Appli	759	6	1.5	1070	4	US-08-697-954-2	Sequence 2, Appli
687	6	1.5	834	5	PCT-US91-07035-8	Sequence 8, Appli	760	6	1.5	1070	4	US-08-778-508-13	Sequence 13, Appli
688	6	1.5	834	5	PCT-US91-07035-10	Sequence 10, Appli	761	6	1.5	1070	4	US-09-059-584-13	Sequence 13, Appli
689	6	1.5	834	5	PCT-US95-14418-2	Sequence 2, Appli	762	6	1.5	1072	4	US-09-877-730-18	Sequence 18, Appli
690	6	1.5	834	5	PCT-US95-15327-2	Sequence 2, Appli	763	6	1.5	1074	3	US-08-613-009A-7	Sequence 7, Appli
691	6	1.5	834	5	PCT-US96-10823-2	Sequence 2, Appli	764	6	1.5	1074	4	US-08-778-508-9	Sequence 9, Appli
692	6	1.5	837	2	US-08-216-260-2	Sequence 2, Appli	765	6	1.5	1074	4	US-09-059-584-9	Sequence 9, Appli
693	6	1.5	870	2	US-09-010-928B-2	Sequence 2, Appli	766	6	1.5	1076	2	US-08-867-941-19	Sequence 19, Appli
694	6	1.5	871	4	US-09-255-829-8	Sequence 8, Appli	767	6	1.5	1076	4	US-09-074-658-19	Sequence 19, Appli
695	6	1.5	871	4	US-09-255-829-8	Sequence 8, Appli	768	6	1.5	1089	1	US-08-375-300-2	Sequence 2, Appli
696	6	1.5	871	4	US-08-255-829-26	Sequence 26, Appli	769	6	1.5	1089	3	PCT-US95-16930-2	Sequence 2, Appli
697	6	1.5	873	4	US-09-255-829-6	Sequence 6, Appli	770	6	1.5	1089	5	PCT-US95-16930-2	Sequence 5, Appli
698	6	1.5	875	4	US-09-255-829-10	Sequence 10, Appli	771	6	1.5	1090	4	US-09-085-198B-5	Sequence 5, Appli
699	6	1.5	878	4	US-09-255-829-12	Sequence 12, Appli	772	6	1.5	1150	4	US-09-877-730-8	Sequence 8, Appli
700	6	1.5	884	4	US-09-741-150-4	Sequence 4, Appli	773	6	1.5	1162	4	US-08-827-962-15	Sequence 15, Appli
701	6	1.5	894	4	US-09-255-829-4	Sequence 4, Appli	774	6	1.5	1162	4	US-08-827-962-20	Sequence 20, Appli
702	6	1.5	904	4	US-09-877-730-6	Sequence 6, Appli	775	6	1.5	1162	4	US-08-803-346-1	Sequence 1, Appli
703	6	1.5	907	4	US-09-255-829-16	Sequence 16, Appli	776	6	1.5	1176	1	US-08-257-999-2	Sequence 2, Appli
704	6	1.5	907	4	US-09-877-730-20	Sequence 20, Appli	777	6	1.5	1213	4	US-09-413-814-79	Sequence 79, Appli
705	6	1.5	953	4	US-09-255-829-14	Sequence 14, Appli	778	6	1.5	1213	4	US-09-134-001C-5319	Sequence 5319, Ap
706	6	1.5	985	4	US-09-877-730-10	Sequence 10, Appli	779	6	1.5	1235	1	US-08-118-101A-2	Sequence 2, Appli
707	6	1.5	991	4	US-09-877-730-12	Sequence 12, Appli	780	6	1.5	1237	1	US-08-241-853-2	Sequence 2, Appli
708	6	1.5	992	1	US-08-127-499A-1	Sequence 1, Appli	781	6	1.5	1237	2	US-08-850-917-2	Sequence 2, Appli
709	6	1.5	992	1	US-08-482-847-1	Sequence 1, Appli	782	6	1.5	1250	3	US-08-938-291A-9	Sequence 9, Appli
710	6	1.5	995	5	PCT-US95-04910-14	Sequence 14, Appli	783	6	1.5	1263	4	US-09-446-504-6	Sequence 6, Appli
711	6	1.5	1012	1	US-08-216-276A-19	Sequence 2, Appli	784	6	1.5	1263	4	US-09-712-266-6	Sequence 6, Appli
712	6	1.5	1012	1	US-07-944-943-2	Sequence 2, Appli	785	6	1.5	1263	4	US-09-091-889A-4	Sequence 4, Appli
713	6	1.5	1012	1	US-07-944-525-2	Sequence 2, Appli	786	6	1.5	1296	2	US-08-480-604A-28	Sequence 28, Appli
714	6	1.5	1012	1	US-08-219-262B-1	Sequence 1, Appli	787	6	1.5	1296	2	US-08-405-496A-28	Sequence 28, Appli
715	6	1.5	1012	1	US-08-219-262B-2	Sequence 2, Appli	788	6	1.5	1296	4	US-08-915-136-28	Sequence 28, Appli
716	6	1.5	1012	1	US-08-219-262B-3	Sequence 3, Appli	789	6	1.5	1447	4	US-09-376-330-17	Sequence 17, Appli
717	6	1.5	1012	1	US-08-219-262B-4	Sequence 4, Appli	790	6	1.5	1503	4	US-08-976-255-14	Sequence 14, Appli
718	6	1.5	1012	1	US-08-219-262B-5	Sequence 5, Appli	791	6	1.5	1536	4	US-09-413-814-10	Sequence 10, Appli
719	6	1.5	1012	1	US-08-219-262B-6	Sequence 6, Appli	792	6	1.5	1548	4	US-09-376-330-15	Sequence 15, Appli
720	6	1.5	1012	1	US-08-219-262B-7	Sequence 7, Appli	793	6	1.5	1614	4	US-09-052-469-2	Sequence 2, Appli
721	6	1.5	1012	1	US-08-219-262B-8	Sequence 8, Appli	794	6	1.5	1626	2	US-08-771-602D-2	Sequence 2, Appli
722	6	1.5	1012	1	US-08-219-262B-9	Sequence 9, Appli	795	6	1.5	1626	4	US-09-232-446B-2	Sequence 2, Appli
723	6	1.5	1012	1	US-08-219-262B-10	Sequence 10, Appli	796	6	1.5	1754	1	US-07-745-206A-13	Sequence 13, Appli
724	6	1.5	1012	1	US-08-219-262B-12	Sequence 12, Appli	797	6	1.5	1754	2	US-08-311-363-13	Sequence 13, Appli
725	6	1.5	1012	1	US-08-219-262B-14	Sequence 14, Appli	798	6	1.5	1872	1	US-08-188-582-14	Sequence 14, Appli
726	6	1.5	1012	2	US-08-708-541A-30	Sequence 30, Appli	799	6	1.5	1872	1	US-08-646-715-14	Sequence 14, Appli
727	6	1.5	1012	3	US-09-031-655-1	Sequence 1, Appli	800	6	1.5	1893	1	US-08-188-582-11	Sequence 11, Appli
728	6	1.5	1012	3	US-09-031-655-2	Sequence 2, Appli	801	6	1.5	1893	1	US-08-646-715-11	Sequence 11, Appli
729	6	1.5	1012	3	US-09-031-655-3	Sequence 3, Appli	802	6	1.5	1996	2	US-08-804-227C-9	Sequence 9, Appli
730	6	1.5	1012	3	US-09-031-655-4	Sequence 4, Appli	803	6	1.5	1996	2	US-08-804-198-3	Sequence 3, Appli
731	6	1.5	1012	3	US-09-031-655-5	Sequence 5, Appli	804	6	1.5	2409	6	5180808-2	Patent No. 5180808
732	6	1.5	1012	3	US-09-031-655-6	Sequence 6, Appli	805	6	1.5	2509	1	US-08-469-005A-10	Sequence 10, Appli
733	6	1.5	1012	3	US-09-031-655-7	Sequence 7, Appli	806	6	1.5	2511	4	US-09-261-907-2	Sequence 2, Appli
734	6	1.5	1012	3	US-09-031-655-8	Sequence 8, Appli	807	6	1.5	2549	4	US-08-471-112A-3	Sequence 3, Appli
735	6	1.5	1012	3	US-09-031-655-9	Sequence 9, Appli	808	6	1.5	2549	5	PCT-US95-06722-12	Sequence 12, Appli
736	6	1.5	1012	3	US-09-031-655-10	Sequence 10, Appli	809	6	1.5	2890	4	US-09-413-814-67	Sequence 67, Appli
737	6	1.5	1012	3	US-09-031-655-12	Sequence 12, Appli	810	6	1.5	2987	2	US-08-970-269A-29	Sequence 29, Appli
738	6	1.5	1012	3	US-09-031-655-14	Sequence 14, Appli	811	6	1.5	2987	4	US-09-407-562-29	Sequence 29, Appli
739	6	1.5	1013	2	US-08-708-541A-34	Sequence 34, Appli	812	6	1.5	3457	2	US-08-416-603-4	Sequence 4, Appli
740	6	1.5	1013	4	US-09-415-522-8	Sequence 8, Appli	813	6	1.5	3798	3	US-09-335-409-6	Sequence 6, Appli
741	6	1.5	1013	4	US-09-255-829-18	Sequence 18, Appli	814	6	1.5	3798	4	US-09-568-102-6	Sequence 6, Appli
742	6	1.5	1022	2	US-08-304-309-2	Sequence 2, Appli	815	6	1.5	3798	4	US-09-567-969-6	Sequence 6, Appli
743	6	1.5	1022	3	US-08-991-942-2	Sequence 2, Appli	816	6	1.5	3798	4	US-09-568-480-6	Sequence 6, Appli
744	6	1.5	1025	3	US-09-138-103-2	Sequence 2, Appli	817	6	1.5	3798	4	US-09-568-486-6	Sequence 6, Appli
745	6	1.5	1025	5	PCT-US95-04567-2	Sequence 2, Appli	818	6	1.5	3798	4	US-09-568-472-6	Sequence 6, Appli
746	6	1.5	1025	5	PCT-US95-04567-4	Sequence 4, Appli	819	6	1.5	3798	4	US-08-567-899-6	Sequence 6, Appli
747	6	1.5	1025	5	PCT-US95-04567-7	Sequence 7, Appli	820	6	1.5	3959	2	US-08-970-269A-30	Sequence 30, Appli
748	6	1.5	1052	4	US-08-778-570B-14	Sequence 14, Appli	821	6	1.5	3959	4	US-09-407-562-30	Sequence 30, Appli
749	6	1.5	1052	4	US-09-059-584-14	Sequence 14, Appli	822	6	1.5	4302	3	US-08-658-136-5	Sequence 5, Appli
750	6	1.5	1053	3	US-08-613-009A-8	Sequence 8, Appli	823	6	1.5	4302	4	US-09-052-469-8	Sequence 8, Appli
751	6	1.5	1053	4	US-08-778-508B-10	Sequence 10, Appli	824	6	1.5	4303	2	US-08-460-751-2	Sequence 2, Appli
752	6	1.5	1053	4	US-09-059-584-10	Sequence 10, Appli	825	6	1.5	4339	4	US-09-052-469-6	Sequence 6, Appli
753	6	1.5	1063	1	US-08-093-453B-3	Sequence 3, Appli	826	6	1.5	5215	5	US-09-105-537-2	Sequence 2, Appli
754	6	1.5	1063	1	US-08-127-499A-8	Sequence 8, Appli	827	5	1.2	5	4	US-09-541-965A-1	Sequence 1, Appli
755	6	1.5	1063	1	US-08-482-847-8	Sequence 8, Appli	828	5	1.2	5	4	US-09-544-791-1	Sequence 1, Appli
756	6	1.5	1067	4	US-09-134-001C-3800	Sequence 3800, Ap	829	5	1.2	6	4	US-08-983-047-3	Sequence 3, Appli
757	6	1.5	1069	4	US-09-877-730-2	Sequence 2, Appli	830	5	1.2	7	3	US-08-411-768B-10	Sequence 10, Appli



831	5	1.2	7	3	US-08-411-768B-13	Sequence 13, Appl1	904	5	1.2	17	1	US-08-077-256-2	Sequence 2, Appl1
832	5	1.2	7	4	US-09-084-605B-1	Sequence 1, Appl1	905	5	1.2	17	1	US-08-370-567-9	Sequence 9, Appl1
833	5	1.2	8	1	US-08-08-181A-4	Sequence 4, Appl1	906	5	1.2	17	1	US-08-438-759-9	Sequence 9, Appl1
834	5	1.2	8	2	US-08-747-137-75	Sequence 25, Appl1	907	5	1.2	17	1	US-08-259-672-2	Sequence 2, Appl1
835	5	1.2	8	3	US-08-785-247-26	Sequence 26, Appl1	908	5	1.2	17	1	US-08-538-911-21	Sequence 21, Appl1
836	5	1.2	8	3	US-09-101-167-17	Sequence 17, Appl1	909	5	1.2	17	1	US-08-459-351-2	Sequence 2, Appl1
837	5	1.2	9	2	US-08-347-335A-7	Sequence 7, Appl1	910	5	1.2	17	1	US-08-142-489-1	Sequence 1, Appl1
838	5	1.2	9	3	US-08-159-339A-1091	Sequence 1091, Ap	911	5	1.2	17	2	US-08-08-533-2	Sequence 2, Appl1
839	5	1.2	9	4	US-08-679-006-26	Sequence 26, Appl1	912	5	1.2	17	3	US-08-351-134-1	Sequence 1, Appl1
840	5	1.2	9	4	US-09-105-839D-28	Sequence 28, Appl1	913	5	1.2	17	4	US-08-132-528-5	Sequence 5, Appl1
841	5	1.2	9	4	US-09-105-839D-62	Sequence 62, Appl1	914	5	1.2	17	4	US-08-873-494-4	Sequence 4, Appl1
842	5	1.2	9	4	US-09-105-839D-71	Sequence 71, Appl1	915	5	1.2	17	4	US-09-599-366-5	Sequence 5, Appl1
843	5	1.2	10	1	US-08-366-591-6	Sequence 6, Appl1	916	5	1.2	17	4	US-08-875-494-4	Sequence 4, Appl1
844	5	1.2	10	2	US-08-554-591-2	Sequence 2, Appl1	917	5	1.2	17	5	PCT-US94-05684-9	Sequence 21, Appl1
845	5	1.2	10	4	US-08-339-853A-20	Sequence 20, Appl1	918	5	1.2	17	5	PCT-US94-05684-9	Sequence 2, Appl1
846	5	1.2	11	4	US-08-918-148-6	Sequence 6, Appl1	919	5	1.2	17	5	PCT-US94-06654-2	Sequence 2, Appl1
847	5	1.2	11	4	US-08-197-793-9	Sequence 9, Appl1	920	5	1.2	18	1	US-07-798-223A-2	Sequence 2, Appl1
848	5	1.2	11	1	US-08-406-347A-17	Sequence 17, Appl1	921	5	1.2	18	2	US-08-978-404B-30	Sequence 30, Appl1
849	5	1.2	11	1	US-08-455-633A-6	Sequence 6, Appl1	922	5	1.2	18	4	US-09-177-249-300	Sequence 300, App
850	5	1.2	11	2	US-08-636-176-9	Sequence 9, Appl1	923	5	1.2	18	4	US-08-990-888-25	Sequence 25, Appl1
851	5	1.2	11	2	US-08-456-460C-6	Sequence 6, Appl1	924	5	1.2	18	4	US-08-990-888-28	Sequence 28, Appl1
852	5	1.2	11	4	US-08-969-046-7	Sequence 7, Appl1	925	5	1.2	18	4	US-08-990-888-62	Sequence 62, Appl1
853	5	1.2	11	5	PCT-US94-05354-6	Sequence 6, Appl1	926	5	1.2	18	4	US-08-990-888-65	Sequence 65, Appl1
854	5	1.2	11	5	PCT-US95-01618-9	Sequence 9, Appl1	927	5	1.2	19	1	US-08-596-983-3	Sequence 3, Appl1
855	5	1.2	12	1	US-07-721-761A-12	Sequence 12, Appl1	928	5	1.2	19	2	US-08-643-034A-7	Sequence 7, Appl1
856	5	1.2	12	1	US-07-978-687-12	Sequence 12, Appl1	929	5	1.2	19	2	US-08-630-961-1	Sequence 1, Appl1
857	5	1.2	12	1	US-08-117-362-28	Sequence 28, Appl1	930	5	1.2	19	3	US-08-648-650A-7	Sequence 7, Appl1
858	5	1.2	12	1	US-08-190-788A-282	Sequence 282, App	931	5	1.2	19	4	US-09-392-055-10	Sequence 10, Appl1
859	5	1.2	12	1	US-08-313-127A-9	Sequence 9, Appl1	932	5	1.2	19	4	US-09-326-976-1	Sequence 1, Appl1
860	5	1.2	12	1	US-08-883-474B-283	Sequence 283, App	933	5	1.2	19	4	US-09-441-502B-11	Sequence 11, Appl1
861	5	1.2	12	1	US-08-465-931A-282	Sequence 282, App	934	5	1.2	20	1	US-08-526-710-17	Sequence 17, Appl1
862	5	1.2	12	1	US-08-486-924-28	Sequence 28, Appl1	935	5	1.2	20	1	US-08-614-935-18	Sequence 18, Appl1
863	5	1.2	12	2	US-08-464-538B-281	Sequence 281, App	936	5	1.2	20	2	US-08-596-387B-81	Sequence 81, Appl1
864	5	1.2	12	4	US-09-307-265A-12	Sequence 12, Appl1	937	5	1.2	20	2	US-08-596-387B-89	Sequence 89, Appl1
865	5	1.2	12	4	US-07-861-458C-143	Sequence 143, App	938	5	1.2	20	3	US-08-862-855-17	Sequence 17, Appl1
866	5	1.2	12	5	PCT-US91-05801-12	Sequence 12, Appl1	939	5	1.2	20	3	US-09-130-287-18	Sequence 18, Appl1
867	5	1.2	13	1	US-08-122-520C-4	Sequence 4, Appl1	940	5	1.2	20	4	US-08-612-973-68	Sequence 68, Appl1
868	5	1.2	13	2	US-08-748-021-32	Sequence 32, Appl1	941	5	1.2	20	4	US-08-927-597-68	Sequence 68, Appl1
869	5	1.2	13	2	US-08-748-021-56	Sequence 56, Appl1	942	5	1.2	20	4	US-09-345-466-4	Sequence 4, Appl1
870	5	1.2	13	3	US-08-974-297-32	Sequence 32, Appl1	943	5	1.2	20	4	US-09-226-985-17	Sequence 17, Appl1
871	5	1.2	13	3	US-08-974-297-56	Sequence 56, Appl1	944	5	1.2	20	4	US-09-227-906-17	Sequence 17, Appl1
872	5	1.2	13	3	US-09-556-601-28	Sequence 28, Appl1	945	5	1.2	20	4	US-09-067-615-81	Sequence 81, Appl1
873	5	1.2	13	5	PCT-US93-07261-14	Sequence 14, Appl1	946	5	1.2	20	4	US-09-067-615-89	Sequence 89, Appl1
874	5	1.2	14	1	US-08-281-193-12	Sequence 12, Appl1	947	5	1.2	20	4	US-09-011-922A-7	Sequence 7, Appl1
875	5	1.2	14	1	US-08-422-106-12	Sequence 12, Appl1	948	5	1.2	20	4	US-09-414-453A-4	Sequence 4, Appl1
876	5	1.2	14	1	US-07-702-771-4	Sequence 4, Appl1	949	5	1.2	20	4	US-09-149-476-542	Sequence 542, App
877	5	1.2	14	1	US-08-279-058B-3	Sequence 3, Appl1	950	5	1.2	20	5	PCT-US95-09816A-81	Sequence 81, Appl1
878	5	1.2	14	1	US-08-438-883-4	Sequence 4, Appl1	951	5	1.2	20	5	PCT-US95-09816A-89	Sequence 89, Appl1
879	5	1.2	14	1	US-08-741-599-4	Sequence 4, Appl1	952	5	1.2	21	2	US-08-347-335A-29	Sequence 29, Appl1
880	5	1.2	14	2	US-08-748-021-27	Sequence 27, Appl1	953	5	1.2	21	4	US-09-149-476-541	Sequence 541, App
881	5	1.2	14	2	US-08-735-716-12	Sequence 12, Appl1	954	5	1.2	22	1	US-08-475-983-30	Sequence 30, Appl1
882	5	1.2	14	2	US-07-963-538B-10	Sequence 10, Appl1	955	5	1.2	22	2	US-08-475-983-30	Sequence 30, Appl1
883	5	1.2	14	2	US-08-347-335A-16	Sequence 16, Appl1	956	5	1.2	22	3	US-08-256-833-30	Sequence 30, Appl1
884	5	1.2	14	2	US-08-845-161A-20	Sequence 20, Appl1	957	5	1.2	22	3	US-08-851-843A-154	Sequence 154, App
885	5	1.2	14	3	US-08-555-568B-12	Sequence 12, Appl1	958	5	1.2	22	4	US-08-974-549A-274	Sequence 274, App
886	5	1.2	14	3	US-08-974-297-27	Sequence 27, Appl1	959	5	1.2	22	4	US-08-854-050-154	Sequence 154, App
887	5	1.2	14	4	US-08-679-006-4	Sequence 4, Appl1	960	5	1.2	22	4	US-09-430-322-154	Sequence 154, App
888	5	1.2	14	4	US-09-370-751-20	Sequence 20, Appl1	961	5	1.2	22	6	5185431-24	Sequence 154, App
889	5	1.2	14	4	US-08-687-590-39	Sequence 39, Appl1	962	5	1.2	24	1	US-08-524-677-11	Sequence 11, Appl1
890	5	1.2	14	4	US-09-519-223-12	Sequence 12, Appl1	963	5	1.2	24	2	US-08-347-335A-1	Sequence 1, Appl1
891	5	1.2	14	4	US-08-828-323-3	Sequence 3, Appl1	964	5	1.2	24	2	US-08-465-491-26	Sequence 26, Appl1
892	5	1.2	15	1	PCT-US95-08069-12	Sequence 12, Appl1	965	5	1.2	24	2	US-08-986-617-26	Sequence 26, Appl1
893	5	1.2	15	1	US-08-455-633A-7	Sequence 7, Appl1	966	5	1.2	24	3	US-08-592-500-22	Sequence 22, Appl1
894	5	1.2	15	2	US-08-456-460C-7	Sequence 7, Appl1	967	5	1.2	24	3	US-08-592-500-31	Sequence 31, Appl1
895	5	1.2	15	4	US-09-476-482-13	Sequence 13, Appl1	968	5	1.2	24	3	US-08-195-006-22	Sequence 22, Appl1
896	5	1.2	15	4	US-09-009-953-67	Sequence 67, Appl1	969	5	1.2	24	3	US-08-195-006-31	Sequence 31, Appl1
897	5	1.2	15	5	PCT-US94-05354-7	Sequence 7, Appl1	970	5	1.2	24	5	PCT-US94-07644A-22	Sequence 22, Appl1
898	5	1.2	15	5	PCT-US95-09261-3	Sequence 3, Appl1	971	5	1.2	24	5	PCT-US94-07644A-31	Sequence 31, Appl1
899	5	1.2	16	4	US-08-642-406A-12	Sequence 12, Appl1	972	5	1.2	25	4	US-09-143-124-13	Sequence 13, Appl1
900	5	1.2	16	4	US-08-505-250-20	Sequence 20, Appl1	973	5	1.2	25	4	US-09-143-124-14	Sequence 14, Appl1
901	5	1.2	16	4	US-08-505-250-20	Sequence 20, Appl1	974	5	1.2	25	4	US-09-143-124-18	Sequence 18, Appl1
902	5	1.2	16	4	US-09-199-534-12	Sequence 12, Appl1	975	5	1.2	25	4	US-08-900-230-38	Sequence 38, Appl1
903	5	1.2	16	4	US-09-199-534-12	Sequence 12, Appl1	976	5	1.2	26	1	US-07-914-284A-4	Sequence 4, Appl1

```

977      5      1.2      26      1      US-07-800-364B-12
978      5      1.2      26      1      US-08-209-261B-12
979      5      1.2      26      2      US-08-596-387B-97
980      5      1.2      26      2      US-08-620-151-92
981      5      1.2      26      3      US-08-928-881-21
982      5      1.2      26      4      US-08-320-424-19
983      5      1.2      26      4      US-09-143-124-15
984      5      1.2      26      4      US-09-143-124-16
985      5      1.2      26      4      US-09-143-124-17
986      5      1.2      26      4      US-09-067-615-97
987      5      1.2      26      4      US-08-264-578-18
988      5      1.2      26      4      US-09-543-921-21
989      5      1.2      26      4      US-09-149-476-532
990      5      1.2      26      5      PCT-US91-03388-12
991      5      1.2      26      5      PCT-US93-06645-4
992      5      1.2      26      5      PCT-US95-09816A-97
993      5      1.2      27      1      US-08-372-783-236
994      5      1.2      27      1      US-08-378-761A-33
995      5      1.2      27      1      US-08-485-286-33
996      5      1.2      27      1      US-08-261-660A-43
997      5      1.2      27      1      US-08-351-365-2
998      5      1.2      27      2      US-08-467-538-2
999      5      1.2      27      3      US-09-119-263-236
1000     5      1.2      27      4      US-09-280-909A-43

```

ALIGNMENTS

```

RESULT 1
US-08-249-013-6
; Sequence 6, Application US/08249013
; Patent No. 5643754
; GENERAL INFORMATION:
; APPLICANT: Haake, David A.
; TITLE OF INVENTION: CLONED leptospira OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Spensley Horn Jubas & Lubitz
; STREET: 1880 Century Park East, Suite 500
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/249,013
; FILING DATE: 25-MAY-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Tumarkin Ph.D., Lisa A.,
; REGISTRATION NUMBER: P-38,347
; REFERENCE/DOCKET NUMBER: PD-3602
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 455-5100
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: Iuta
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..117
; US-08-249-013-6

```

```

Query Match      2.5%; Score 10; DB 1; Length 117;
Best Local Similarity 100.0%; Pred. No. 0.036;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 TGGLINIVTK 163
DB 60 TGGLINIVTK 69
|||||

RESULT 2
US-08-886-863-6
; Sequence 6, Application US/0886863
; Patent No. 5824321
; GENERAL INFORMATION:
; APPLICANT: Haake, David A.
; TITLE OF INVENTION: CLONED leptospira OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Spensley Horn Jubas & Lubitz
; STREET: 1880 Century Park East, Suite 500
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/886,863
; FILING DATE: 01-JUL-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/249,013
; FILING DATE: 25-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Tumarkin Ph.D., Lisa A.,
; REGISTRATION NUMBER: P-38,347
; REFERENCE/DOCKET NUMBER: PD-3602
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 455-5100
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: Iuta
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..117
; US-08-886-863-6

Query Match      2.5%; Score 10; DB 2; Length 117;
Best Local Similarity 100.0%; Pred. No. 0.036;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 TGGLINIVTK 163
DB 60 TGGLINIVTK 69
|||||

RESULT 3
US-09-175-229-6
; Sequence 6, Application US/09175229
; Patent No. 6309641
; GENERAL INFORMATION:
; APPLICANT: Haake, David A.

```

```

; TITLE OF INVENTION: CLONED leptospira OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Spensley Horn Jubb & Lubitz
; STREET: 1880 Century Park East, Suite 500
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/175,229
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/249,013
; FILING DATE: 25-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Tumarkin Ph.D., Lisa A.,
; REGISTRATION NUMBER: P-38,347
; REFERENCE/DOCKET NUMBER: PD-3602
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 455-5100
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: luca
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..117
; US-09-175-229-6

Query Match      2.5%; Score 10; DB 4; Length 117;
Best Local Similarity 100.0%; Pred. No. 0.036;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 TGGIINIYTK 163
|||
DB 60 TGGIINIYTK 69

RESULT 4
PCT-US95-06764-6
; Sequence 6, Application PC/TUS9506764
; GENERAL INFORMATION:
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: CLONED leptospira OUTER MEMBRANE PROTEIN
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06764
; FILING DATE: 25-MAY-1995
```

```

; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.,
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: PD3602
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 678-5070
; TELEFAX: (619) 678-5099
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: luca
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..117
; PCT-US95-06764-6

Query Match      2.5%; Score 10; DB 5; Length 117;
Best Local Similarity 100.0%; Pred. No. 0.036;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 TGGIINIYTK 163
|||
DB 60 TGGIINIYTK 69

RESULT 5
US-08-679-493A-187
; Sequence 187, Application US/08679493A
; Patent No. 6303295
; GENERAL INFORMATION:
; APPLICANT: Taylor, Ethan W.,
; TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS
; FILE REFERENCE: 55-95
; CURRENT APPLICATION NUMBER: US/08/679,493A
; PRIOR FILING DATE: 1996-07-12
; PRIOR APPLICATION NUMBER: 60/001203
; PRIOR FILING DATE: 1995-07-14
; PRIOR APPLICATION NUMBER: 60/003,112
; PRIOR FILING DATE: 1995-09-01
; NUMBER OF SEQ ID NOS: 216
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 187
; LENGTH: 114
; TYPE: PRT
; ORGANISM: rabbit
; US-08-679-493A-187

Query Match      1.8%; Score 7; DB 4; Length 114;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 358 PSATLDA 364
|||
DB 29 PSATLDA 35

RESULT 6
US-08-556-965-2
; Sequence 2, Application US/08556965
; Patent No. 6025540
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Production of EC-SOD
; NUMBER OF SEQUENCES: 31
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/556.965  
FILING DATE:  
CLASSIFICATION: 800  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 222 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-556-965-2

Query Match 1.8%; Score 7; DB 3; Length 222;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 358 PSATLDA 364  
| | | | |  
Db 49 PSATLDA 55

RESULT 7  
5472691-3  
; Patent No. 5472691  
; APPLICANT: MARKLUND, STEFAN; EDLUND, THOMAS  
; TITLE OF INVENTION: SUPEROXIDE DISMUTASE  
; NUMBER OF SEQUENCES: 7  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/125.744  
; FILING DATE: 24-SP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 897.624  
; FILING DATE: 12-JUN-1992  
; APPLICATION NUMBER: 576.114  
; FILING DATE: 27-aug-1990  
; APPLICATION NUMBER: 902.596  
; FILING DATE: 02-sep-1986  
; SEQ ID NO: 3  
; LENGTH: 222  
5472691-3

Query Match 1.8%; Score 7; DB 6; Length 222;  
Best Local Similarity 100.0%; Pred. No. 63;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 358 PSATLDA 364  
| | | | |  
Db 49 PSATLDA 55

RESULT 8  
US-09-134-001C-4827  
; Sequence 4827, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134.001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064.964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055.779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 4827  
; LENGTH: 235  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4827

Query Match 1.8%; Score 7; DB 4; Length 235;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 205 LARLDVD 211  
| | | | |  
Db 80 LARLDVD 86

RESULT 9  
US-08-023-980B-45  
; Sequence 45, Application US/08023980B  
; Patent No. 5843641  
; GENERAL INFORMATION:  
; APPLICANT: Brown, Robert  
; APPLICANT: Horvitz, H. Robert  
; APPLICANT: Rosen, Daniel R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS,  
; TITLE OF INVENTION: TREATMENT AND PREVENTION OF DISEASES OF CELL DEATH  
; NUMBER OF SEQUENCES: 45  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Clark & Elbing LLP  
; STREET: 585 Commercial Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02109-1024  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/023.980B  
; FILING DATE: 26-FEB-1993  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 00786/177001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/723-4123  
; TELEFAX: 617/723-8962  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 45:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 240 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-023-980B-45

Query Match 1.8%; Score 7; DB 2; Length 240;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 358 PSATLDA 364  
| | | | |  
Db 67 PSATLDA 73

RESULT 10  
US-08-486-953A-53  
; Sequence 53, Application US/08486953A  
; Patent No. 5849290  
; GENERAL INFORMATION:  
; APPLICANT: Brown, Robert  
; APPLICANT: Horvitz, H. Robert  
; APPLICANT: Rosen, Daniel R.  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS,  
; TITLE OF INVENTION: TREATMENT AND PREVENTION OF DISEASES OF CELL DEATH  
; NUMBER OF SEQUENCES: 53

;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Clark & Elding LLP  
;; STREET: 176 Federal Street  
;; CITY: Boston  
;; STATE: MA  
;; COUNTRY: USA  
;; ZIP: 02110  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: FastSeq  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/486,953A  
;; FILING DATE: 07-JUN-1995  
;; CLASSIFICATION: 424  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/204,052  
;; FILING DATE: 28-FEB-1994  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Clark, Paul T.  
;; REGISTRATION NUMBER: 30,162  
;; REFERENCE/DOCKET NUMBER: 00786/223002  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 617/428-0200  
;; TELEFAX: 617/428-7045  
;; TELEX:  
;; INFORMATION FOR SEQ ID NO: 53:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 240 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
;; US-08-486-953A-53

Query Match 1.8%; Score 7; DB 2; Length 240;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 358 PSATLDA 364  
DB 67 PSATLDA 73  
RESULT 11  
US-08-679-493A-186  
; Sequence 186, Application US/08679493A  
; Patent No. 6303295  
; GENERAL INFORMATION:  
; APPLICANT: Taylor, Behan W.  
; TITLE OF INVENTION: SELENOPROTEINS, CODING SEQUENCES AND METHODS  
; FILE REFERENCE: 55-95  
; CURRENT APPLICATION NUMBER: US/08/679,493A  
; CURRENT FILING DATE: 1996-07-12  
; PRIOR APPLICATION NUMBER: 60/001203  
; PRIOR FILING DATE: 1995-07-14  
; PRIOR APPLICATION NUMBER: 60/003,112  
; PRIOR FILING DATE: 1995-09-01  
; NUMBER OF SEQ ID NOS: 216  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 186  
; LENGTH: 240  
; TYPE: PRT  
; ORGANISM: bloodfluke  
; US-08-679-493A-186

Query Match 1.8%; Score 7; DB 4; Length 240;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 358 PSATLDA 364  
DB 67 PSATLDA 73

RESULT 12  
5472691-2  
; Patent No. 5472691  
; APPLICANT: MARKLUND, STEFAN; EDLUND, THOMAS  
; TITLE OF INVENTION: SUPEROXIDE DISMUTASE  
; NUMBER OF SEQUENCES: 7  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/125,744  
; FILING DATE: 24-SEP-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 897,624  
; FILING DATE: 12-JUN-1992  
; APPLICATION NUMBER: 576,114  
; FILING DATE: 27-AUG-1990  
; APPLICATION NUMBER: 902,596  
; FILING DATE: 02-SEP-1986  
; SEQ ID NO: 2  
; LENGTH: 240  
5472691-2

Query Match 1.8%; Score 7; DB 6; Length 240;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 358 PSATLDA 364  
DB 67 PSATLDA 73

RESULT 13  
US-09-134-001C-3317  
; Sequence 3317, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 3317  
; LENGTH: 298  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
; US-09-134-001C-3317

Query Match 1.8%; Score 7; DB 4; Length 298;  
Best Local Similarity 100.0%; Pred. No. 84;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 290 LNAIKGL 296  
DB 251 LNAIKGL 257

RESULT 14  
US-08-109-391A-2  
; Sequence 2, Application US/08109391A  
; Patent No. 5639876  
; GENERAL INFORMATION:  
; APPLICANT: Tripp, Cynthia A.  
; APPLICANT: Frank, Glenn R.  
; APPLICANT: Grievie, Robert B.  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING NOVEL  
; TITLE OF INVENTION: PARASITIC HELMINTH PROTEINS  
; NUMBER OF SEQUENCES: 17

```
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/109,391A
; FILING DATE: 19-AUG-1993
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-13
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/863-9700
; TELEFAX: 303/863-0223
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-109-391A-2

Query Match 1.8%; Score 7; DB 1; Length 303;
Best Local Similarity 100.0%; Pred.No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 65 YEEQIQE 71
Db 253 YEEQIQE 259

RESULT 15
US-08-459-019A-2
; Sequence 2, Application US/08459019A
; Patent No. 5686080
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Frank, Glenn R.
; APPLICANT: Grieve, Robert B.
; TITLE OF INVENTION: NOVEL PARASITIC HELMINTH P4 PROTEINS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, #3500
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,019A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 33,020
; REFERENCE/DOCKET NUMBER: 2618-13-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
```

```
;
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-459-019A-2

Query Match 1.8%; Score 7; DB 1; Length 303;
Best Local Similarity 100.0%; Pred.No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 65 YEEQIQE 71
Db 253 YEEQIQE 259

RESULT 16
US-08-460-428A-2
; Sequence 2, Application US/08460428A
; Patent No. 5912337
; GENERAL INFORMATION:
; APPLICANT: Tripp, Cynthia A.
; APPLICANT: Frank, Glenn R.
; APPLICANT: Grieve, Robert B.
; TITLE OF INVENTION: NOVEL PARASITIC HELMINTH
; TITLE OF INVENTION: P22U PROTEINS
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,428A
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-13-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303/863-9700
; TELEFAX: 303/863-0223
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-460-428A-2

Query Match 1.8%; Score 7; DB 2; Length 303;
Best Local Similarity 100.0%; Pred.No. 85;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 65 YEEQIQE 71
Db 253 YEEQIQE 259

RESULT 17
US-08-458-860A-2
; Sequence 2, Application US/08458860A
; Patent No. 6100390
; GENERAL INFORMATION:
```

APPLICANT: Frank, Glenn R.  
APPLICANT: Tripp, Cynthia A.  
APPLICANT: Grievé, Robert B.  
TITLE OF INVENTION: NOVEL PARASITIC HELMINTH  
TITLE OF INVENTION: P22U NUCLEIC ACID MOLECULES  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sheridan Ross P.C.  
STREET: 1700 Lincoln St., Suite 3500  
CITY: Denver  
STATE: CO  
COUNTRY: U.S.A.  
ZIP: 80203  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,860A  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Connell, Gary J.  
REGISTRATION NUMBER: 32,020  
REFERENCE/DOCKET NUMBER: 2618-13-2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 303/863-9700  
TELEFAX: 303/863-0223  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 303 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-458-860A-2

Query Match 1.8%; Score 7; DB 3; Length 303;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 YEEQIOE 71  
Db 253 YEEQIOE 259

RESULT 18  
US-08-229-287-4  
Sequence 4, Application US/08229287  
Patent No. 5530193  
GENERAL INFORMATION:  
APPLICANT: Clark Jr., John M.  
APPLICANT: Ulka, Joseph M.  
APPLICANT: Murry, Lynn B.  
APPLICANT: Scarafia, Lilliana E.  
TITLE OF INVENTION: VIRUS RESISTANT CORN PLANTS  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sandoz Agro, Inc.  
STREET: 975 California Avenue  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/229,287  
FILING DATE:  
TOPOLOGY: linear  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/067,257  
FILING DATE: 25-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/817,922  
FILING DATE: 08-JAN-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Marcus-Wyner, Lynn  
REGISTRATION NUMBER: 34,869  
REFERENCE/DOCKET NUMBER: 135-1084/XCC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/354-3588  
TELEFAX: 415/857-1125  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 328 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-229-287-4

Query Match 1.8%; Score 7; DB 1; Length 328;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 GSGATCG 156  
Db 59 GSGATCG 65

RESULT 19  
US-08-153-848-44  
Sequence 44, Application US/08153848  
Patent No. 5759804  
GENERAL INFORMATION:  
APPLICANT: Godiska, Ronald  
APPLICANT: Gray, Patrick W.  
APPLICANT: Schweikart, Vicki L.  
TITLE OF INVENTION: No. 5759804e1 Seven Transmembrane Receptors  
NUMBER OF SEQUENCES: 64  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
ADDRESSEE: Bicknell  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/153,848  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/977,452  
FILING DATE: 17-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 5759804and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 31794  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 44:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 339 amino acids  
TYPE: amino acid  
TOPOLOGY: linear

MOLECULE TYPE: protein  
US-08-153-848-44

Query Match 1.8%; Score 7; DB 1; Length 339;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 ALANRIT 55  
Db 276 ALANRIT 282

## RESULT 20

US-08-812-871-3  
Sequence 3, Application US/08812871  
Patent No. 5955303

## GENERAL INFORMATION:

APPLICANT: Au-Young, Janice  
APPLICANT: Guegler, Karl  
APPLICANT: Muzong Cheng  
TITLE OF INVENTION: NOVEL HUMAN CHEMOKINE RECEPTOR-LIKE  
TITLE OF INVENTION: PROTEIN  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Dr.  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/812,871

FILING DATE: Filed Herewith  
PRIOR APPLICATION DATA:

APPLICATION NUMBER:  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J

REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0237 US  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:

LENGTH: 339 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:

LIBRARY: GenBank  
CLONE: 992700

US-08-812-871-3

Query Match 1.8%; Score 7; DB 2; Length 339;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 ALANRIT 55  
Db 276 ALANRIT 282

## RESULT 21

US-09-299-843A-44  
Sequence 44, Application US/09299843A  
Patent No. 6107475

## GENERAL INFORMATION:

APPLICANT: Godiska, Ronald  
APPLICANT: Gray, Patrick W.  
APPLICANT: Schweikart, Vicki L.  
TITLE OF INVENTION: No. 6107475el Seven Transmembrane Receptors  
NUMBER OF SEQUENCES: 66  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
ADDRESSEE: Borun  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/299,843A  
FILING DATE:

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 09/088,337  
FILING DATE: 01-JUN-1998

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/153,848

FILING DATE: 17-NOV-1993  
APPLICATION NUMBER: US 07/977,452

PRIOR APPLICATION DATA:  
FILING DATE: 17-NOV-1992

ATTORNEY/AGENT INFORMATION:  
NAME: Jill E. Uhl

REGISTRATION NUMBER: 43,213  
REFERENCE/DOCKET NUMBER: 27866/32059B

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300

TELEFAX: (312) 474-0448  
TELEX:

INFORMATION FOR SEQ ID NO: 44:  
SEQUENCE CHARACTERISTICS:

LENGTH: 339 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: protein

US-09-299-843A-44

Query Match 1.8%; Score 7; DB 3; Length 339;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 ALANRIT 55  
Db 276 ALANRIT 282

## RESULT 22

US-09-088-337B-44  
Sequence 44, Application US/09088337B  
Patent No. 6348574

## GENERAL INFORMATION:

APPLICANT: Godiska, Ronald  
Gray, Patrick W.

Schweikart, Vicki L.

TITLE OF INVENTION: No. 6348574el Seven Transmembrane Receptors  
NUMBER OF SEQUENCES: 86

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
Borun

STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois



COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/088,337B  
FILING DATE: 01-Jun-1998  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/153,848  
FILING DATE: 17-NOV-1993  
APPLICATION NUMBER: US 07/977,452  
FILING DATE: 17-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: No. 6348574and, Greta E.  
REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 31794  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 44:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 339 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 44:  
US-09-088-337B-44

Query Match 1.8%; Score 7; DB 4; Length 339;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 ALANRIT 55  
Db 276 ALANRIT 282

RESULT 23  
PCT-US93-11153-44  
Sequence 44, Application PC/TUS9311153  
GENERAL INFORMATION:  
APPLICANT: Godiska, Ronald  
APPLICANT: Gray, Patrick W.  
APPLICANT: Schwellhart, Vicki L.  
TITLE OF INVENTION: Novel Seven Transmembrane Receptors  
NUMBER OF SEQUENCES: 64  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
STREET: 6300 Sears Tower, 233 South Wacker Drive  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/11153  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/977,452  
FILING DATE: 17-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Notand, Greta E.

REGISTRATION NUMBER: 35,302  
REFERENCE/DOCKET NUMBER: 31794  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 474-6300  
TELEFAX: (312) 474-0448  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 44:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 339 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US93-11153-44

Query Match 1.8%; Score 7; DB 5; Length 339;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 ALANRIT 55  
Db 276 ALANRIT 282

RESULT 24  
PCT-US95-07180-2  
Sequence 2, Application PC/TUS9507180  
GENERAL INFORMATION:  
APPLICANT: LI, YI  
APPLICANT: GOCAYNE, JEANINE D  
APPLICANT: RUBEN, STEVEN M  
TITLE OF INVENTION: G-PROTEIN RECEPTOR HIBB69  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,  
STREET: 6 BECKER FARM ROAD  
CITY: ROSELAND  
STATE: NJ  
COUNTRY: US  
ZIP: 07068  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/07180  
FILING DATE: 06-JUN-1995  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: MULLINS, J.G.  
REGISTRATION NUMBER: 30,073  
REFERENCE/DOCKET NUMBER: 325800-366  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201-994-1700  
TELEFAX: 201-994-1744  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 339 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-07180-2

Query Match 1.8%; Score 7; DB 5; Length 339;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 49 ALANRIT 55  
Db 276 ALANRIT 282

```
RESULT 25
US-09-416-050A-4
; Sequence 4, Application US/09416050A
; Patent No. 6194559
; GENERAL INFORMATION:
; APPLICANT: KIM, Soo Young
; TITLE OF INVENTION: Abscisic Acid Responsive Element -Binding Transcription Factors
; FILE REFERENCE: 1942/42
; CURRENT APPLICATION NUMBER: US/09/416.050A
; CURRENT FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-416-050A-4

Query Match      1.8%; Score 7; DB 4; Length 416;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 LSSEGIG 188
Db 298 LSSEGIG 304
|||||

RESULT 26
US-09-664-800-4
; Sequence 4, Application US/09664800
; Patent No. 6218527
; GENERAL INFORMATION:
; APPLICANT: KIM, Soo Young
; TITLE OF INVENTION: Abscisic Acid Responsive Element -Binding Transcription Factors
; FILE REFERENCE: 1942/42
; CURRENT APPLICATION NUMBER: US/09/664.800
; CURRENT FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: 09/416.050
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-664-800-4

Query Match      1.8%; Score 7; DB 4; Length 416;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 LSSEGIG 188
Db 298 LSSEGIG 304
|||||

RESULT 27
US-09-665-309-4
; Sequence 4, Application US/09665309
; Patent No. 6232461
; GENERAL INFORMATION:
; APPLICANT: KIM, Soo Young
; TITLE OF INVENTION: Abscisic Acid Responsive Element -Binding Transcription Factors
; FILE REFERENCE: 1942/42
; CURRENT APPLICATION NUMBER: US/09/665.309
; CURRENT FILING DATE: 2000-09-19
; PRIOR APPLICATION NUMBER: 09/416.050
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-665-309-4

Query Match      1.8%; Score 7; DB 4; Length 416;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 LSSEGIG 188
Db 298 LSSEGIG 304
|||||
```

```
; ORGANISM: Arabidopsis thaliana
US-09-665-309-4

Query Match      1.8%; Score 7; DB 4; Length 416;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 LSSEGIG 188
Db 298 LSSEGIG 304
|||||

RESULT 28
US-09-661-569-4
; Sequence 4, Application US/09661569
; Patent No. 6245905
; GENERAL INFORMATION:
; APPLICANT: KIM, Soo Young
; TITLE OF INVENTION: Abscisic Acid Responsive Element -Binding Transcription Factors
; FILE REFERENCE: 1942/42
; CURRENT APPLICATION NUMBER: US/09/661.569
; CURRENT FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 09/416.050
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-661-569-4

Query Match      1.8%; Score 7; DB 4; Length 416;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 LSSEGIG 188
Db 298 LSSEGIG 304
|||||

RESULT 29
US-08-676-444-40
; Sequence 40, Application US/08676444A
; Patent No. 6294325
; GENERAL INFORMATION:
; APPLICANT: Wetmur, James G.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF THERMOSTABLE
; TITLE OF INVENTION: MUTL GENES AND PROTEINS AND USES THEREFOR
; FILE REFERENCE: MSM95-02
; CURRENT APPLICATION NUMBER: US/08/676.444A
; CURRENT FILING DATE: 1996-07-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 40
; LENGTH: 426
; TYPE: PRT
; ORGANISM: Aquifex pyrophilus
US-08-676-444-40

Query Match      1.8%; Score 7; DB 4; Length 426;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 141 VLSGATS 147
Db 72 VLSGATS 78
|||||

RESULT 30
US-08-496-944-2
; Sequence 2, Application US/08496944
; Patent No. 6040496
```

GENERAL INFORMATION:  
APPLICANT: Law, Marcus D  
TITLE OF INVENTION: Use of Translationally altered RNA to  
TITLE OF INVENTION: Confer Resistance to Maize Dwarf Mosaic Virus and Other  
TITLE OF INVENTION: Monocyledonous Plant Viruses  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-Geigy Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: NY  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30B  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/496,944  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Eimet, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: CGC 1814  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2763 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-496-944-2

Query Match 1.8%; Score 7; DB 3; Length 2763;  
Best Local Similarity 100.0%; Pred. No. 7.7e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 150 GSGATGG 156  
Db 2494 GSGATGG 2500

RESULT 31  
US-09-041-886-43  
Sequence 43, Application US/09041886  
GENERAL INFORMATION:  
APPLICANT: Bredesen, Dale E.  
TITLE OF INVENTION: Proapoptotic Peptides, Dependence  
TITLE OF INVENTION: Polypeptides and Methods of Use  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/041,886  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815

REFERENCE/DOCKET NUMBER: P-LJ 2626  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 11 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-041-886-43

Query Match 1.5%; Score 6; DB 4; Length 11;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 359 SATLDA 364  
Db 1 SATLDA 6

RESULT 32  
US-07-841-997A-42  
Sequence 42, Application US/07841997A  
Patent No. 5422254  
GENERAL INFORMATION:  
APPLICANT: Vucorio, John  
TITLE OF INVENTION: A method to increase the trehalose content  
TITLE OF INVENTION: of organisms by transforming them with the  
TITLE OF INVENTION: structural genes for the short and long chains  
TITLE OF INVENTION: yeast trehalose synthase.  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Alko Ltd.  
STREET: PO Box 350  
CITY: Helsinki  
STATE: Finland  
COUNTRY: Finland  
ZIP: SF-00101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb  
COMPUTER: IBM PC/XT/AT  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: MP5.1 file exported as DOS text file  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/841,997A  
FILING DATE: 19920228  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/836,021  
FILING DATE: February 14, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Mary E. Gormley  
REGISTRATION NUMBER: 34409  
REFERENCE/DOCKET NUMBER: 920085A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)659-2930  
TELEFAX: (202)887-0357  
TELEX: 440142  
INFORMATION FOR SEQ ID NO: 42:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 12 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: Peptide  
HYPOTHETICAL: NO  
FRAGMENT TYPE: N-terminal  
US-07-841-997A-42

Query Match 1.5%; Score 6; DB 1; Length 12;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 GATGGL 157  
|||||  
Db 3 GATGGL 8

RESULT 33  
US-08-290-301-42  
; Sequence 42, Application US/08290301  
; Patent No. 5792921  
; GENERAL INFORMATION:  
; APPLICANT: Londerborough, John  
; APPLICANT: Tunnele, Outi  
; APPLICANT: Palva, Tupio  
; APPLICANT: Holmstrom, Kjell-Ove  
; APPLICANT: Welin, Bjorn  
; APPLICANT: Mandel, Abul  
; TITLE OF INVENTION: Increasing the trehalose content  
; TITLE OF INVENTION: of organisms by transforming them with combinations of  
; TITLE OF INVENTION: the structural genes for trehalose synthase.  
; NUMBER OF SEQUENCES: 85  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Alko Ltd.  
; STREET: PO Box 350  
; CITY: Helsinki  
; STATE: -  
; COUNTRY: Finland  
; ZIP: SF-00101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb  
; COMPUTER: IBM PC/XT/AT  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: WPS.1 file exported as DOS text file  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/290,301  
; FILING DATE: 15 August 1994  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: FI 943133  
; FILING DATE: 29 June 1994  
; APPLICATION NUMBER: PCT/FI93/00049  
; FILING DATE: 15 February 1993  
; APPLICATION NUMBER: 07/841,997  
; FILING DATE: 28 February 1992  
; APPLICATION NUMBER: 07/836,021  
; FILING DATE: 14 February 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kubovcik, Ronald J.  
; NAME: Lydon, James C.  
; REGISTRATION NUMBER: 25,401  
; REGISTRATION NUMBER: 30,082  
; REFERENCE/DOCKET NUMBER: LAIN-001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 467-6300  
; TELEFAX: (202) 466-2006  
; INFORMATION FOR SEQ ID NO: 42:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12 amino acids  
; TYPE: Amino acid  
; TOPOLOGY: Linear  
; MOLECULE TYPE: Peptide  
; HYPOTHETICAL: NO  
; FRAGMENT TYPE: N-terminal  
US-08-290-301-42

Query Match 1.5%; Score 6; DB 1; Length 12;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 GATGGL 157  
|||||  
Db 3 GATGGL 8

RESULT 34  
US-09-013-598-42  
; Sequence 42, Application US/09013598  
; Patent No. 6323001  
; GENERAL INFORMATION:  
; APPLICANT: Londerborough, John  
; APPLICANT: Tunnele, Outi  
; APPLICANT: Palva, Tupio  
; APPLICANT: Holmstrom, Kjell-Ove  
; APPLICANT: Welin, Bjorn  
; APPLICANT: Mandel, Abul  
; TITLE OF INVENTION: Increasing the trehalose content  
; TITLE OF INVENTION: of organisms by transforming them with combinations of  
; TITLE OF INVENTION: the structural genes for trehalose synthase.  
; NUMBER OF SEQUENCES: 85  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Alko Ltd.  
; STREET: PO Box 350  
; CITY: Helsinki  
; STATE: -  
; COUNTRY: Finland  
; ZIP: SF-00101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb  
; COMPUTER: IBM PC/XT/AT  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: WPS.1 file exported as DOS text file  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/013,598  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/290,301  
; FILING DATE:  
; APPLICATION NUMBER: PCT/FI93/00049  
; FILING DATE: 15 February 1993  
; APPLICATION NUMBER: 07/841,997  
; FILING DATE: 28 February 1992  
; APPLICATION NUMBER: 07/836,021  
; FILING DATE: 14 February 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kubovcik, Ronald J.  
; NAME: Lydon, James C.  
; REGISTRATION NUMBER: 25,401  
; REGISTRATION NUMBER: 30,082  
; REFERENCE/DOCKET NUMBER: LAIN-001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 467-6300  
; TELEFAX: (202) 466-2006  
; INFORMATION FOR SEQ ID NO: 42:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12 amino acids  
; TYPE: Amino acid  
; TOPOLOGY: Linear  
; MOLECULE TYPE: Peptide  
; HYPOTHETICAL: No  
; FRAGMENT TYPE: N-terminal  
US-09-013-598-42

Query Match 1.5%; Score 6; DB 4; Length 12;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 152 GATGGL 157  
|||||  
Db 3 GATGGL 8

RESULT 35  
US-09-041-886-3  
; Sequence 3, Application US/09041886  
; Patent No. 6235872

GENERAL INFORMATION:  
APPLICANT: Bredesen, Dale E.  
APPLICANT: Rabizadeh, Sharriz  
TITLE OF INVENTION: Proapoptotic Peptides, Dependence  
TITLE OF INVENTION: Polypeptides and Methods of Use  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/041,886  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LJ 2626  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-041-886-3

Query Match 1.5%; Score 6; DB 4; Length 14;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 SATLDA 364  
|||||  
DB 1 SATLDA 6

RESULT 36  
US-09-041-886-4  
Sequence 4, Application US/09041886  
Patent No. 6235872  
GENERAL INFORMATION:  
APPLICANT: Bredesen, Dale E.  
APPLICANT: Rabizadeh, Sharriz  
TITLE OF INVENTION: Proapoptotic Peptides, Dependence  
TITLE OF INVENTION: Polypeptides and Methods of Use  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/041,886  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LJ 2626  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-041-886-4

Query Match 1.5%; Score 6; DB 4; Length 14;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 SATLDA 364  
|||||  
DB 1 SATLDA 6

RESULT 37  
US-09-041-886-5  
Sequence 5, Application US/09041886  
Patent No. 6235872  
GENERAL INFORMATION:  
APPLICANT: Bredesen, Dale E.  
APPLICANT: Rabizadeh, Sharriz  
TITLE OF INVENTION: Proapoptotic Peptides, Dependence  
TITLE OF INVENTION: Polypeptides and Methods of Use  
NUMBER OF SEQUENCES: 72  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: United States  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/041,886  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LJ 2626  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 14 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-041-886-5

Query Match 1.5%; Score 6; DB 4; Length 14;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 SATLDA 364  
|||||  
DB 1 SATLDA 6

RESULT 38  
US-09-041-886-41  
; Sequence 41, Application US/09041886  
; Patent No. 6235872  
; GENERAL INFORMATION:  
; APPLICANT: Bredesen, Dale E.  
; APPLICANT: Rabizadeh, Sharro  
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence  
; TITLE OF INVENTION: Polypeptides and Methods of Use  
; NUMBER OF SEQUENCES: 72  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell & Flores LLP  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: United States  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/041.886  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; REFERENCE/DOCKET NUMBER: P-LJ 2626  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 41:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 14 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-041-886-41  
  
Query Match 1.5%; Score 6; DB 4; Length 14;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 359 SATLDA 364  
Db 1 SATLDA 6  
|||||  
  
RESULT 39  
US-08-554-612C-45  
; Sequence 45, Application US/08554612C  
; Patent No. 5747660  
; GENERAL INFORMATION:  
; APPLICANT: Orlicky, David  
; TITLE OF INVENTION: PROSTAGLANDIN F2' RECEPTOR REGULATORY  
; TITLE OF INVENTION: PROTEIN AND THERAPEUTIC USES  
; NUMBER OF SEQUENCES: 51  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 2200 Sand Hill Road, Suite 100  
; CITY: Menlo Park  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 94025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/554.612C

; FILING DATE: No. 5747660ember 6, 1995  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sherwood, Pamela  
; REGISTRATION NUMBER: 36,677  
; REFERENCE/DOCKET NUMBER: 06519/004001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 322-5070  
; TELEFAX: (415) 854-0875  
; INFORMATION FOR SEQ ID NO: 45:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 17 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-554-612C-45  
  
Query Match 1.5%; Score 6; DB 1; Length 17;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 14 LLSVAV 19  
Db 9 LLSVAV 14  
|||||  
  
RESULT 40  
US-09-549-831-7  
; Sequence 7, Application US/09549831  
; Patent No. 6429305  
; GENERAL INFORMATION:  
; APPLICANT: Chang, Chi-Yao  
; APPLICANT: Chang, Chia-Ching  
; APPLICANT: Leu, Kuen-Lin  
; APPLICANT: Tsai, Chih-Tung  
; APPLICANT: Ting, Jing-Wen  
; APPLICANT: Lin, Chih-Hung  
; TITLE OF INVENTION: FISH GROWTH HORMONES  
; FILE REFERENCE: 08191-039001  
; CURRENT APPLICATION NUMBER: US/09/549.831  
; CURRENT FILING DATE: 2000-04-14  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 19  
; TYPE: PRT  
; ORGANISM: Epinephelus awoara  
US-09-549-831-7  
  
Query Match 1.5%; Score 6; DB 4; Length 19;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 89 SLGVSS 94  
Db 12 SLGVSS 17  
|||||  
  
RESULT 41  
US-08-256-747C-9  
; Sequence 9, Application US/08256747C  
; Patent No. 6037448  
; GENERAL INFORMATION:  
; APPLICANT: CHONG, pele  
; APPLICANT: GILLAM, Shirley  
; APPLICANT: OU, Dawei  
; APPLICANT: TINGLE, Aubrey  
; TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR A RUBELLA VACCINE  
; NUMBER OF SEQUENCES: 78  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor 701, 330 University Avenue  
; CITY: Toronto

```

STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/256,747C
FILING DATE: 06-OCT-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-370
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-256-747C-9

Query Match
Best Local Similarity 1.5%; Score 6; DB 3; Length 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 SVAGVS 199
DB 14 SVAGVS 19

RESULT 42
US-08-834-130A-9
Sequence 9, Application US/08834130A
Patent No. 6180758
GENERAL INFORMATION:
APPLICANT: CHONG, Pele
APPLICANT: GILLAM, Shirley
APPLICANT: OU, Dawei
APPLICANT: TINGLE, Aubrey
TITLE OF INVENTION: SYNTHETIC PEPTIDES FOR A RUBELLA VACCINE
NUMBER OF SEQUENCES: 75
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/834,130A
FILING DATE: 14-APR-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: STEWART, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-686 MIS:jb
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids

```

```

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-834-130A-9

Query Match
Best Local Similarity 1.5%; Score 6; DB 4; Length 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 SVAGVS 199
DB 14 SVAGVS 19

RESULT 43
US-08-295-411-6
Sequence 6, Application US/08295411
Patent No. 5679639
GENERAL INFORMATION:
APPLICANT: Griffin, John H.
APPLICANT: Meesters, Rolf M.
TITLE OF INVENTION: Serine Protease-Derived Polypeptides and
TITLE OF INVENTION: Anti-Peptide Antibodies, Systems and Therapeutic Methods
TITLE OF INVENTION: for Inhibiting Coagulation
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Office of Patent Counsel, The Scripps
ADDRESSEE: Research Institute
STREET: 10666 No. 5679639th Torrey Pines Road, TPC 8
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/295,411
FILING DATE: 22-AUG-1994
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/793,989
FILING DATE: 18-NOV-1991
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: TSI1263.0C1
TELEPHONE: 619-554-2937
TELEFAX: 619-554-6312
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
US-08-295-411-6

Query Match
Best Local Similarity 1.5%; Score 6; DB 1; Length 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 SLPPLS 16
DB 12 SLPPLS 17

```

```

RESULT 44
US-08-955-471-6
; Sequence 6, Application US/08955471
; Patent No. 5968751
; GENERAL INFORMATION:
; APPLICANT: Griffin, John H.
; APPLICANT: Mesters, Rolf M.
; TITLE OF INVENTION: Serine Protease-Derived Polypeptides and
; TITLE OF INVENTION: Anti-Peptide Antibodies, Systems and Therapeutic Methods
; TITLE OF INVENTION: for Inhibiting Coagulation
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Office of Patent Counsel, The Scripps
; ADDRESSEE: Research Institute
; STREET: 10666 No. 5968751th Torrey Pines Road, TPC 8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/955.471
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/295,411
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: TSRI263.0C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; PCT-US-955-471-6

Query Match 1.5%; Score 6; DB 2; Length 21;
Best Local Similarity 100.0%; Pred.No. 58;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 SLPLLS 16
Db 12 SLPLLS 17

Search completed: December 25, 2002, 20:33:14
Job time : 32 secs

RESULT 45
PCT-US92-10242-6
; Sequence 6, Application PC/TUS9210242
; GENERAL INFORMATION:
; APPLICANT: Griffin, John H.
; APPLICANT: Mesters, Rolf
; TITLE OF INVENTION: Serine Protease-Derived Polypeptides and
; TITLE OF INVENTION: Anti-Peptide Antibodies, Systems and Therapeutic Methods
; TITLE OF INVENTION: for Inhibiting Coagulation
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Office of Patent Counsel, The Scripps

```

```

; ADDRESSEE: Research Institute
; STREET: 10666 North Torrey Pines Road, TPC 8
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/10242
; FILING DATE: 19921118
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/793,989
; FILING DATE: 18-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitting, Thomas
; REGISTRATION NUMBER: 34,163
; REFERENCE/DOCKET NUMBER: SCRO472P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-554-2937
; TELEFAX: 619-554-6312
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; PCT-US92-10242-6

Query Match 1.5%; Score 6; DB 5; Length 21;
Best Local Similarity 100.0%; Pred.No. 58;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 SLPLLS 16
Db 12 SLPLLS 17

Search completed: December 25, 2002, 20:33:14
Job time : 32 secs

```



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compen Ltd.

OM protein - protein search, using sw model

Run on: December 25, 2002, 20:32:08 ; Search time 12 Seconds

(without alignments)  
577.138 Million cell updates/sec

Title: US-09-889-746-2

Perfect score: 400  
Sequence: 1 MKHSHPQWLSLPLSVAVT.....SKAEVLGRVINKPKRALF 400

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 109717 seqs, 1731436 residues

Word size : 0

Total number of hits satisfying chosen parameters: 109717

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :

Published Applications\_AA:\*  
1: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/2/pubppaa/PTCT\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB pep.\*  
5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/2/pubppaa/PTCTUS\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB pep.\*  
9: /cgn2\_6/ptodata/2/pubppaa/US09\_NEW\_PUB pep.\*  
10: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB pep.\*  
11: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB pep.\*  
12: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB pep.\*  
13: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB pep.\*  
14: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	2.5	117	12	US-10-022-461-6
2	9	2.2	813	10	US-09-815-242-5085
3	7	1.8	285	10	US-09-815-242-5436
4	7	1.8	293	10	US-09-815-242-12145
5	7	1.8	293	10	US-09-815-242-12799
6	7	1.8	300	10	US-09-393-634-17
7	7	1.8	305	9	US-10-008-118A-8
8	7	1.8	305	10	US-09-443-704-8
9	7	1.8	332	9	US-10-042-141-72
10	7	1.8	332	10	US-09-726-643-72
11	7	1.8	339	9	US-09-828-478-4
12	7	1.8	339	10	US-09-848-889-12
13	7	1.8	339	10	US-09-788-133-2
14	7	1.8	367	9	US-09-828-478-6
15	7	1.8	386	10	US-09-881-752A-74
16	7	1.8	453	9	US-10-042-141-46
17	7	1.8	453	10	US-09-726-643-46
18	7	1.8	489	9	US-10-042-141-82
19	7	1.8	489	10	US-09-726-643-82

20	7	1.8	489	10	US-09-925-300-1151	Sequence 1151, App
21	7	1.8	648	10	US-09-741-669-330	Sequence 330, App
22	7	1.8	648	10	US-09-815-242-10089	Sequence 10089, A
23	6	1.5	24	9	US-09-915-789A-13	Sequence 13, Appl
24	6	1.5	26	9	US-10-016-634A-104	Sequence 104, App
25	6	1.5	37	10	US-09-864-761-48223	Sequence 48223, A
26	6	1.5	40	10	US-09-864-761-47372	Sequence 47372, A
27	6	1.5	42	10	US-09-864-761-33872	Sequence 33872, A
28	6	1.5	42	10	US-09-864-761-40375	Sequence 40375, A
29	6	1.5	45	10	US-09-995-494-94	Sequence 94, Appl
30	6	1.5	47	10	US-09-864-761-42399	Sequence 42399, A
31	6	1.5	50	10	US-09-864-761-39663	Sequence 39663, A
32	6	1.5	50	10	US-09-864-761-45308	Sequence 45308, A
33	6	1.5	56	10	US-09-864-761-34894	Sequence 34894, A
34	6	1.5	56	10	US-09-779-451-76	Sequence 76, Appl
35	6	1.5	58	12	US-10-035-408-4	Sequence 4, Appl
36	6	1.5	58	10	US-09-864-847-525	Sequence 525, App
37	6	1.5	70	10	US-09-867-550-1698	Sequence 1698, App
38	6	1.5	84	10	US-09-764-877-1603	Sequence 1603, App
39	6	1.5	86	10	US-09-864-761-34118	Sequence 34118, A
40	6	1.5	92	9	US-09-982-405-2	Sequence 2, Appl
41	6	1.5	92	9	US-10-001-054-20	Sequence 20, Appl
42	6	1.5	92	12	US-10-032-586-446	Sequence 446, App
43	6	1.5	93	9	US-09-981-876-203	Sequence 203, App
44	6	1.5	93	9	US-09-981-876-256	Sequence 256, App
45	6	1.5	95	10	US-09-864-761-48467	Sequence 48467, A
46	6	1.5	101	10	US-09-925-299-830	Sequence 830, App
47	6	1.5	102	10	US-09-764-877-1421	Sequence 1421, App
48	6	1.5	107	10	US-09-939-960-486	Sequence 486, App
49	6	1.5	117	10	US-09-821-831-10	Sequence 10, Appl
50	6	1.5	119	10	US-09-925-299-799	Sequence 799, App
51	6	1.5	120	10	US-09-925-297-505	Sequence 505, App
52	6	1.5	121	9	US-09-764-866-838	Sequence 838, App
53	6	1.5	124	9	US-09-992-598-266	Sequence 266, App
54	6	1.5	124	9	US-09-989-293A-266	Sequence 266, App
55	6	1.5	124	9	US-09-989-735-266	Sequence 266, App
56	6	1.5	124	9	US-09-990-444-266	Sequence 266, App
57	6	1.5	124	10	US-09-989-722-266	Sequence 266, App
58	6	1.5	124	10	US-09-989-723-266	Sequence 266, App
59	6	1.5	124	10	US-09-989-729-266	Sequence 266, App
60	6	1.5	124	10	US-09-989-727-266	Sequence 266, App
61	6	1.5	124	10	US-09-989-731-266	Sequence 266, App
62	6	1.5	124	10	US-09-989-732-266	Sequence 266, App
63	6	1.5	124	10	US-09-991-073-266	Sequence 266, App
64	6	1.5	124	10	US-09-990-442-266	Sequence 266, App
65	6	1.5	124	10	US-09-991-163-266	Sequence 266, App
66	6	1.5	124	10	US-09-993-604-266	Sequence 266, App
67	6	1.5	124	10	US-09-990-456-266	Sequence 266, App
68	6	1.5	124	10	US-09-989-721-266	Sequence 266, App
69	6	1.5	130	10	US-09-925-301-864	Sequence 864, App
70	6	1.5	130	10	US-09-764-853-538	Sequence 538, App
71	6	1.5	132	10	US-09-864-761-33532	Sequence 33532, A
72	6	1.5	143	12	US-10-052-586-440	Sequence 340, App
73	6	1.5	144	10	US-09-925-299-985	Sequence 985, App
74	6	1.5	144	10	US-09-815-242-5693	Sequence 5693, App
75	6	1.5	144	10	US-09-815-242-12260	Sequence 12260, A
76	6	1.5	144	10	US-09-815-242-12763	Sequence 12763, A
77	6	1.5	147	10	US-09-841-132-519	Sequence 519, App
78	6	1.5	150	10	US-09-746-339A-66	Sequence 66, Appl
79	6	1.5	153	10	US-09-929-300-1631	Sequence 1631, App
80	6	1.5	154	10	US-09-909-745-6	Sequence 6, Appl
81	6	1.5	157	10	US-09-864-761-44299	Sequence 44299, A
82	6	1.5	168	10	US-09-864-018-115	Sequence 115, App
83	6	1.5	170	10	US-09-909-745-8	Sequence 8, Appl
84	6	1.5	173	10	US-09-272-809-16	Sequence 16, Appl
85	6	1.5	178	10	US-09-909-745-2	Sequence 2, Appl
86	6	1.5	189	10	US-09-987-967-2	Sequence 2, Appl
87	6	1.5	191	9	US-09-991-496-82	Sequence 82, Appl
88	6	1.5	191	10	US-09-874-993-82	Sequence 82, Appl
89	6	1.5	194	10	US-09-777-422-14	Sequence 14, Appl
90	6	1.5	199	10	US-09-949-476-2	Sequence 2, Appl
91	6	1.5	199	10	US-09-949-271-2	Sequence 2, Appl
92	6	1.5	199	10	US-09-949-272-2	Sequence 2, Appl

93	6	1.5	199	10	US-09-777-422-2	Sequence 2, Appli	166	6	1.5	415	10	US-09-815-242-12099	Sequence 12099, A
94	6	1.5	207	10	US-09-746-359A-65	Sequence 55, Appli	167	6	1.5	416	10	US-09-805-848A-2	Sequence 2, Appli
95	6	1.5	208	10	US-09-815-242-5302	Sequence 5302, Ap	168	6	1.5	416	10	US-09-805-848A-4	Sequence 4, Appli
96	6	1.5	214	10	US-09-746-359A-63	Sequence 63, Appli	169	6	1.5	419	10	US-09-894-018-123	Sequence 123, App
97	6	1.5	214	10	US-09-924-256A-32	Sequence 32, Appli	170	6	1.5	423	10	US-09-073-009-142	Sequence 142, App
98	6	1.5	215	9	US-09-915-789A-20	Sequence 20, Appli	171	6	1.5	423	10	US-09-793-306-142	Sequence 142, App
99	6	1.5	215	10	US-09-909-745-14	Sequence 14, Appli	172	6	1.5	425	10	US-09-748-537-14	Sequence 14, Appli
100	6	1.5	216	10	US-09-789-561-158	Sequence 158, App	173	6	1.5	425	10	US-09-821-831-2	Sequence 2, Appli
101	6	1.5	217	10	US-09-746-359A-55	Sequence 55, Appli	174	6	1.5	427	8	US-08-681-219-26	Sequence 26, Appli
102	6	1.5	219	10	US-09-815-242-12168	Sequence 12168, A	175	6	1.5	427	10	US-09-826-213-5	Sequence 5, Appli
103	6	1.5	219	10	US-09-815-242-13043	Sequence 13043, A	176	6	1.5	427	10	US-09-748-537-13	Sequence 13, Appli
104	6	1.5	220	10	US-09-925-301-1209	Sequence 1209, Ap	177	6	1.5	427	10	US-09-935-727-7	Sequence 7, Appli
105	6	1.5	221	10	US-09-746-359A-12	Sequence 12, Appli	178	6	1.5	429	10	US-09-815-242-14022	Sequence 14022, A
106	6	1.5	221	10	US-09-925-302-607	Sequence 607, App	179	6	1.5	432	10	US-09-815-243-13058	Sequence 13058, A
107	6	1.5	222	10	US-09-893-737-108	Sequence 108, App	180	6	1.5	433	10	US-09-815-242-13781	Sequence 13781, A
108	6	1.5	224	9	US-10-108-605-277	Sequence 277, App	181	6	1.5	433	10	US-09-801-368-38	Sequence 38, Appli
109	6	1.5	226	10	US-09-909-745-4	Sequence 4, Appli	182	6	1.5	436	9	US-10-011-588-27	Sequence 27, Appli
110	6	1.5	231	10	US-09-815-242-12671	Sequence 12671, A	183	6	1.5	436	12	US-10-078-929-64	Sequence 64, Appli
111	6	1.5	235	10	US-09-747-116-4	Sequence 4, Appli	184	6	1.5	437	9	US-10-216-441-4	Sequence 4, Appli
112	6	1.5	247	9	US-09-792-793A-39	Sequence 39, Appli	185	6	1.5	437	10	US-09-818-656A-4	Sequence 4, Appli
113	6	1.5	251	10	US-09-821-831-4	Sequence 4, Appli	186	6	1.5	441	10	US-09-925-301-1303	Sequence 1303, Ap
114	6	1.5	287	10	US-09-864-761-42836	Sequence 42836, A	187	6	1.5	443	9	US-09-895-913A-150	Sequence 150, App
115	6	1.5	287	10	US-09-323-998D-17	Sequence 17, Appli	188	6	1.5	444	9	US-10-011-588-9	Sequence 9, Appli
116	6	1.5	299	10	US-09-886-055-257	Sequence 257, App	189	6	1.5	447	10	US-09-916-109-5	Sequence 5, Appli
117	6	1.5	310	9	US-09-934-352-2	Sequence 2, Appli	190	6	1.5	448	10	US-09-288-326-7	Sequence 7, Appli
118	6	1.5	314	10	US-09-771-161A-189	Sequence 189, App	191	6	1.5	449	9	US-10-011-588-47	Sequence 47, Appli
119	6	1.5	314	10	US-09-771-161A-190	Sequence 190, App	192	6	1.5	450	10	US-09-896-720-2	Sequence 2, Appli
120	6	1.5	314	10	US-09-771-161A-191	Sequence 191, App	193	6	1.5	450	10	US-09-765-111A-14	Sequence 14, Appli
121	6	1.5	315	10	US-09-910-174A-28	Sequence 28, Appli	194	6	1.5	450	10	US-09-765-111A-25	Sequence 25, Appli
122	6	1.5	315	10	US-09-886-055-125	Sequence 125, App	195	6	1.5	451	10	US-09-862-027-45	Sequence 45, Appli
123	6	1.5	316	9	US-09-978-285A-137	Sequence 137, App	196	6	1.5	454	10	US-09-815-242-5209	Sequence 5209, Ap
124	6	1.5	316	9	US-09-896-738-14	Sequence 14, Appli	197	6	1.5	459	9	US-10-216-441-2	Sequence 2, Appli
125	6	1.5	316	9	US-09-915-789A-1	Sequence 1, Appli	198	6	1.5	459	10	US-09-818-656A-2	Sequence 2, Appli
126	6	1.5	316	9	US-09-915-789A-3	Sequence 3, Appli	199	6	1.5	460	10	US-09-815-242-10945	Sequence 10945, A
127	6	1.5	316	9	US-09-978-697-137	Sequence 137, App	200	6	1.5	461	9	US-10-011-588-5	Sequence 5, Appli
128	6	1.5	316	9	US-09-978-192A-137	Sequence 137, App	201	6	1.5	463	12	US-10-037-860-13	Sequence 13, Appli
129	6	1.5	316	9	US-09-998-832A-137	Sequence 137, App	202	6	1.5	465	10	US-09-933-561-15	Sequence 15, Appli
130	6	1.5	316	10	US-09-789-561-135	Sequence 135, App	203	6	1.5	468	10	US-09-815-243-5144	Sequence 5144, Ap
131	6	1.5	316	10	US-09-875-338-11	Sequence 11, Appli	204	6	1.5	474	10	US-09-737-149-34	Sequence 34, Appli
132	6	1.5	316	10	US-09-875-338-13	Sequence 13, Appli	205	6	1.5	474	10	US-09-737-149-37	Sequence 37, Appli
133	6	1.5	316	10	US-09-910-174A-24	Sequence 24, Appli	206	6	1.5	475	10	US-09-815-242-10052	Sequence 10052, A
134	6	1.5	316	10	US-09-955-866-8	Sequence 8, Appli	207	6	1.5	479	8	US-08-910-386A-24	Sequence 24, Appli
135	6	1.5	316	12	US-10-052-586-54	Sequence 54, Appli	208	6	1.5	483	10	US-09-916-109-4	Sequence 4, Appli
136	6	1.5	320	8	US-08-914-350-2	Sequence 2, Appli	209	6	1.5	490	10	US-09-815-242-10057	Sequence 10057, A
137	6	1.5	320	10	US-09-789-777-28	Sequence 28, Appli	210	6	1.5	499	10	US-09-905-999-21	Sequence 21, Appli
138	6	1.5	321	10	US-09-970-711-22	Sequence 22, Appli	211	6	1.5	505	12	US-10-135-689-2	Sequence 2, Appli
139	6	1.5	324	10	US-09-886-055-217	Sequence 217, App	212	6	1.5	505	12	US-10-135-689-4	Sequence 4, Appli
140	6	1.5	338	9	US-09-730-763-4	Sequence 4, Appli	213	6	1.5	522	10	US-09-815-242-11002	Sequence 11002, A
141	6	1.5	338	9	US-09-730-763-37	Sequence 37, Appli	214	6	1.5	523	10	US-09-753-008-2	Sequence 2, Appli
142	6	1.5	351	10	US-09-916-109-7	Sequence 7, Appli	215	6	1.5	525	9	US-09-935-371-23	Sequence 23, Appli
143	6	1.5	353	10	US-09-815-242-11564	Sequence 11564, A	216	6	1.5	525	9	US-10-044-722-3	Sequence 3, Appli
144	6	1.5	357	10	US-09-815-242-12009	Sequence 12009, A	217	6	1.5	534	10	US-09-770-643A-14	Sequence 14, Appli
145	6	1.5	358	10	US-09-925-302-493	Sequence 493, App	218	6	1.5	534	10	US-09-875-338-7	Sequence 7, Appli
146	6	1.5	359	10	US-09-881-752A-248	Sequence 248, App	219	6	1.5	541	10	US-09-815-242-11316	Sequence 11316, A
147	6	1.5	359	10	US-09-994-427A-11	Sequence 11, Appli	220	6	1.5	542	12	US-10-052-586-398	Sequence 398, App
148	6	1.5	362	10	US-09-908-419-6	Sequence 6, Appli	221	6	1.5	547	10	US-09-746-359A-54	Sequence 54, Appli
149	6	1.5	370	9	US-09-712-363-187	Sequence 187, App	222	6	1.5	549	9	US-10-051-902-30	Sequence 30, Appli
150	6	1.5	370	9	US-09-870-759-74	Sequence 74, Appli	223	6	1.5	552	9	US-09-935-371-27	Sequence 27, Appli
151	6	1.5	380	9	US-10-160-669-2	Sequence 2, Appli	224	6	1.5	553	9	US-09-935-371-25	Sequence 25, Appli
152	6	1.5	384	10	US-09-896-055-87	Sequence 87, Appli	225	6	1.5	553	9	US-09-935-371-29	Sequence 29, Appli
153	6	1.5	386	10	US-09-815-242-10427	Sequence 10427, A	226	6	1.5	553	10	US-09-746-359A-11	Sequence 11, Appli
154	6	1.5	387	10	US-09-789-561-156	Sequence 156, App	227	6	1.5	553	10	US-09-949-192-7	Sequence 7, Appli
155	6	1.5	387	10	US-09-916-109-6	Sequence 6, Appli	228	6	1.5	559	10	US-09-746-359A-62	Sequence 62, Appli
156	6	1.5	389	10	US-09-816-028A-34	Sequence 34, Appli	229	6	1.5	561	10	US-09-922-138-8	Sequence 8, Appli
157	6	1.5	390	10	US-09-925-297-531	Sequence 531, App	230	6	1.5	562	10	US-09-741-669-363	Sequence 363, App
158	6	1.5	392	9	US-10-094-080-1	Sequence 1, Appli	231	6	1.5	571	10	US-09-746-359A-53	Sequence 53, Appli
159	6	1.5	394	10	US-09-815-242-10233	Sequence 10233, A	232	6	1.5	572	10	US-09-815-242-5536	Sequence 5526, Ap
160	6	1.5	394	10	US-09-863-475A-4	Sequence 4, Appli	233	6	1.5	572	10	US-09-815-242-12351	Sequence 12351, A
161	6	1.5	394	10	US-09-946-034-2	Sequence 2, Appli	234	6	1.5	582	10	US-09-770-643A-12	Sequence 12, Appli
162	6	1.5	411	10	US-09-815-242-5875	Sequence 5875, Ap	235	6	1.5	590	10	US-09-815-242-10812	Sequence 10812, A
163	6	1.5	414	9	US-10-063-547-102	Sequence 102, App	236	6	1.5	591	10	US-09-815-242-5837	Sequence 5837, Ap
164	6	1.5	414	12	US-10-006-867-102	Sequence 102, App	237	6	1.5	594	10	US-09-746-359A-23	Sequence 23, Appli
165	6	1.5	414	12	US-10-052-586-314	Sequence 314, App	238	6	1.5	599	9	US-09-771-382-6	Sequence 6, Appli

239	6	1.5	599	10	US-09-797-862-15	Sequence 15, Appl	312	6	1.5	998	9	US-10-106-534-2	Sequence 2, Appl
240	6	1.5	605	9	US-09-906-209-4	Sequence 4, Appl	313	6	1.5	1003	9	US-09-992-558-33	Sequence 33, Appl
241	6	1.5	607	10	US-09-815-242-13379	Sequence 13379, A	314	6	1.5	1003	9	US-09-989-293A-33	Sequence 33, Appl
242	6	1.5	607	10	US-09-815-242-13682	Sequence 13682, A	315	6	1.5	1003	9	US-09-989-735-33	Sequence 33, Appl
243	6	1.5	612	10	US-09-925-301-1358	Sequence 1358, Ap	316	6	1.5	1003	9	US-09-990-444-33	Sequence 33, Appl
244	6	1.5	612	10	US-09-815-242-11088	Sequence 11088, A	317	6	1.5	1003	10	US-09-989-722-33	Sequence 33, Appl
245	6	1.5	615	10	US-09-899-482-3	Sequence 12, Appl	318	6	1.5	1003	10	US-09-989-723-33	Sequence 33, Appl
246	6	1.5	617	10	US-09-817-676A-12	Sequence 13, Appl	319	6	1.5	1003	10	US-09-989-729-33	Sequence 33, Appl
247	6	1.5	617	10	US-09-815-242-5762	Sequence 5762, Ap	320	6	1.5	1003	10	US-09-989-727-33	Sequence 33, Appl
248	6	1.5	617	10	US-09-863-027-44	Sequence 44, Appl	321	6	1.5	1003	10	US-09-989-731-33	Sequence 33, Appl
249	6	1.5	618	10	US-09-817-676A-14	Sequence 14, Appl	322	6	1.5	1003	10	US-09-989-732-33	Sequence 33, Appl
250	6	1.5	618	10	US-09-970-516-4	Sequence 4, Appl	323	6	1.5	1003	10	US-09-991-073-33	Sequence 33, Appl
251	6	1.5	619	12	US-10-007-693-77	Sequence 77, Appl	324	6	1.5	1003	10	US-09-990-442-33	Sequence 33, Appl
252	6	1.5	635	9	US-10-003-392-14	Sequence 14, Appl	325	6	1.5	1003	10	US-09-991-163-33	Sequence 33, Appl
253	6	1.5	635	10	US-09-925-299-896	Sequence 896, App	326	6	1.5	1003	10	US-09-993-604-33	Sequence 33, Appl
254	6	1.5	636	10	US-09-815-242-12061	Sequence 12061, A	327	6	1.5	1003	10	US-09-990-456-33	Sequence 33, Appl
255	6	1.5	638	10	US-09-863-027-41	Sequence 41, Appl	328	6	1.5	1003	10	US-09-989-721-33	Sequence 33, Appl
256	6	1.5	639	9	US-09-906-209-2	Sequence 2, Appl	329	6	1.5	1025	10	US-09-884-886-2	Sequence 2, Appl
257	6	1.5	635	9	US-09-906-209-10	Sequence 10, Appl	330	6	1.5	1043	10	US-09-737-149-8	Sequence 8, Appl
258	6	1.5	661	10	US-09-801-368-422	Sequence 422, App	331	6	1.5	1057	10	US-09-815-242-5798	Sequence 5798, Ap
259	6	1.5	677	10	US-09-745-763-168	Sequence 168, App	332	6	1.5	1096	9	US-09-960-226-4	Sequence 4, Appl
260	6	1.5	677	10	US-09-925-300-1626	Sequence 1626, Ap	333	6	1.5	1107	10	US-09-815-242-12815	Sequence 12815, A
261	6	1.5	682	10	US-09-815-242-11452	Sequence 11452, A	334	6	1.5	1138	12	US-10-036-328A-6	Sequence 6, Appl
262	6	1.5	687	9	US-09-969-384-16	Sequence 16, Appl	335	6	1.5	1175	10	US-09-770-633A-26	Sequence 26, Appl
263	6	1.5	687	9	US-09-969-384-27	Sequence 27, Appl	336	6	1.5	1175	10	US-09-770-633A-30	Sequence 30, Appl
264	6	1.5	697	10	US-09-770-643A-18	Sequence 18, Appl	337	6	1.5	1198	10	US-09-815-242-12446	Sequence 12446, A
265	6	1.5	698	9	US-09-875-338-9	Sequence 9, Appl	338	6	1.5	1210	12	US-10-036-328A-2	Sequence 2, Appl
266	6	1.5	699	9	US-10-008-355-8	Sequence 8, Appl	339	6	1.5	1258	10	US-09-992-543-1	Sequence 1, Appl
267	6	1.5	701	10	US-09-815-242-13002	Sequence 13002, A	340	6	1.5	1259	10	US-09-770-633A-4	Sequence 4, Appl
268	6	1.5	708	10	US-09-925-301-1012	Sequence 1012, Ap	341	6	1.5	1263	10	US-09-971-309-6	Sequence 6, Appl
269	6	1.5	710	10	US-09-287-849-16	Sequence 16, Appl	342	6	1.5	1295	10	US-09-726-949A-1	Sequence 1, Appl
270	6	1.5	712	9	US-10-008-355-2	Sequence 2, Appl	343	6	1.5	1298	10	US-09-770-633A-24	Sequence 24, Appl
271	6	1.5	714	9	US-09-803-589-7	Sequence 7, Appl	344	6	1.5	1298	10	US-09-770-633A-28	Sequence 28, Appl
272	6	1.5	716	9	US-10-001-054-30	Sequence 30, Appl	345	6	1.5	1307	10	US-09-770-633A-2	Sequence 2, Appl
273	6	1.5	727	10	US-09-445-023A-12	Sequence 12, Appl	346	6	1.5	1385	10	US-09-827-998-16	Sequence 16, Appl
274	6	1.5	733	10	US-09-862-027-42	Sequence 42, Appl	347	6	1.5	1620	10	US-09-815-242-1126	Sequence 5126, Ap
275	6	1.5	734	9	US-10-008-355-5	Sequence 5, Appl	348	6	1.5	1770	10	US-09-827-998-3	Sequence 3, Appl
276	6	1.5	739	9	US-10-036-492-13	Sequence 13, Appl	349	6	1.5	1791	10	US-09-827-998-3	Sequence 3, Appl
277	6	1.5	745	10	US-09-770-643A-16	Sequence 16, Appl	350	6	1.5	2008	9	US-09-736-969A-105	Sequence 105, App
278	6	1.5	768	9	US-09-764-868-728	Sequence 728, App	351	6	1.5	2008	10	US-09-736-969A-2	Sequence 2, Appl
279	6	1.5	777	10	US-09-765-111A-2	Sequence 2, Appl	352	6	1.5	2008	10	US-09-736-969A-91	Sequence 91, Appl
280	6	1.5	791	10	US-09-770-643A-42	Sequence 22, Appl	353	6	1.5	2008	10	US-09-736-960-88	Sequence 88, Appl
281	6	1.5	792	10	US-09-815-242-5724	Sequence 5724, Ap	354	6	1.5	2044	10	US-09-815-242-12713	Sequence 12713, A
282	6	1.5	794	10	US-09-815-242-5697	Sequence 5697, Ap	355	6	1.5	3014	10	US-09-737-149-2	Sequence 2, Appl
283	6	1.5	796	10	US-09-801-368-30	Sequence 30, Appl	356	6	1.5	3798	9	US-10-014-717-6	Sequence 6, Appl
284	6	1.5	802	10	US-09-815-242-12668	Sequence 12668, A	357	6	1.5	4999	9	US-09-976-059-14	Sequence 14, Appl
285	6	1.5	811	9	US-10-011-588-29	Sequence 29, Appl	358	6	1.5	5179	9	US-10-025-380-1068	Sequence 1068, Ap
286	6	1.5	811	10	US-09-765-111A-12	Sequence 12, Appl	359	6	1.5	5179	10	US-09-922-217-1068	Sequence 1068, Ap
287	6	1.5	822	10	US-09-826-312-12	Sequence 12, Appl	360	6	1.5	5179	10	US-09-833-263-33	Sequence 33, Appl
288	6	1.5	826	9	US-09-736-457-330	Sequence 330, App	361	6	1.5	5215	9	US-09-860-846-2	Sequence 2, Appl
289	6	1.5	826	9	US-09-902-941-330	Sequence 330, App	362	6	1.5	5215	10	US-09-861-289-2	Sequence 2, Appl
290	6	1.5	826	9	US-10-101-812-10	Sequence 10, Appl	363	6	1.5	5215	10	US-09-818-656A-11	Sequence 11, Appl
291	6	1.5	826	10	US-09-922-958-4	Sequence 4, Appl	364	6	1.5	5215	7	US-09-996-286-117	Sequence 137, App
292	6	1.5	826	10	US-09-833-790-235	Sequence 235, App	365	6	1.5	5215	7	US-09-996-286-155	Sequence 165, App
293	6	1.5	826	12	US-10-028-158-23	Sequence 23, Appl	366	6	1.5	5215	7	US-09-996-288-167	Sequence 167, Appl
294	6	1.5	834	9	US-10-033-297-6	Sequence 6, Appl	367	6	1.5	5215	7	US-09-947-137-1	Sequence 1, Appl
295	6	1.5	836	10	US-09-777-430A-8	Sequence 8, Appl	368	6	1.5	5215	9	US-09-834-765-115	Sequence 115, App
296	6	1.5	836	10	US-09-777-430A-11	Sequence 11, Appl	369	6	1.5	5215	9	US-09-834-765-335	Sequence 235, App
297	6	1.5	838	10	US-09-894-998-45	Sequence 45, Appl	370	6	1.5	5215	9	US-09-834-765-453	Sequence 453, App
298	6	1.5	839	10	US-09-770-643A-60	Sequence 20, Appl	371	6	1.5	5215	9	US-09-834-765-538	Sequence 538, App
299	6	1.5	840	10	US-09-765-111A-4	Sequence 4, Appl	372	6	1.5	5215	9	US-09-894-018-442	Sequence 242, App
300	6	1.5	842	10	US-09-777-430A-15	Sequence 15, Appl	373	6	1.5	5215	10	US-09-834-765-181	Sequence 181, App
301	6	1.5	842	10	US-09-777-430A-30	Sequence 20, Appl	374	6	1.5	5215	10	US-09-834-765-110	Sequence 410, App
302	6	1.5	842	10	US-09-777-430A-23	Sequence 23, Appl	375	6	1.5	5215	10	US-09-834-765-483	Sequence 483, App
303	6	1.5	842	10	US-09-777-430A-26	Sequence 26, Appl	376	6	1.5	5215	10	US-09-834-765-594	Sequence 594, App
304	6	1.5	856	10	US-09-287-849-12	Sequence 12, Appl	377	6	1.5	5215	10	US-09-834-765-603	Sequence 603, App
305	6	1.5	861	9	US-10-011-588-21	Sequence 21, Appl	378	6	1.5	5215	10	US-09-834-765-596	Sequence 696, App
306	6	1.5	874	10	US-09-815-242-11127	Sequence 11127, A	379	6	1.5	5215	14	US-09-927-180-12	Sequence 12, Appl
307	6	1.5	874	10	US-09-765-111A-6	Sequence 6, Appl	380	6	1.5	5215	17	US-09-850-988B-10	Sequence 10, Appl
308	6	1.5	874	10	US-09-737-149-33	Sequence 33, Appl	381	6	1.5	5215	17	US-09-968-561A-110	Sequence 110, App
309	6	1.5	950	10	US-09-321-987B-4	Sequence 4, Appl	382	6	1.5	5215	17	US-09-864-761-40248	Sequence 40248, A
310	6	1.5	965	10	US-09-737-149-4	Sequence 4, Appl	383	6	1.5	5215	17	US-09-864-761-46244	Sequence 46244, A
311	6	1.5	971	10	US-09-737-149-6	Sequence 6, Appl	384	6	1.5	5215	17	US-09-801-852A-12	Sequence 12, Appl

385	5	1.2	17	10	US-09-192-854-68	Sequence 68, Appl	458	5	1.2	38	10	US-09-864-761-33403	Sequence 33403, A
386	5	1.2	18	10	US-09-071-838-300	Sequence 300, Appl	459	5	1.2	38	10	US-09-864-761-38941	Sequence 38941, A
387	5	1.2	19	10	US-09-893-615-45	Sequence 45, Appl	460	5	1.2	38	10	US-09-925-299-1307	Sequence 1307, App
388	5	1.2	19	10	US-09-893-615-47	Sequence 47, Appl	461	5	1.2	39	9	US-09-984-245-144	Sequence 144, App
389	5	1.2	19	10	US-09-893-615-51	Sequence 51, Appl	462	5	1.2	39	10	US-09-864-761-41872	Sequence 41872, A
390	5	1.2	19	12	US-10-036-444-5	Sequence 5, Appl	463	5	1.2	40	10	US-09-843-845-13	Sequence 13, Appl
391	5	1.2	20	9	US-09-973-025-68	Sequence 68, Appl	464	5	1.2	40	10	US-09-864-761-35023	Sequence 35023, A
392	5	1.2	20	10	US-09-832-312-4	Sequence 4, Appl	465	5	1.2	40	10	US-09-864-761-38035	Sequence 38035, A
393	5	1.2	20	10	US-09-848-164-81	Sequence 81, Appl	466	5	1.2	40	10	US-09-864-761-45435	Sequence 45435, A
394	5	1.2	20	10	US-09-848-164-89	Sequence 89, Appl	467	5	1.2	41	10	US-09-843-845-18	Sequence 18, Appl
395	5	1.2	20	10	US-09-864-761-40075	Sequence 40075, A	468	5	1.2	41	10	US-09-864-761-33473	Sequence 33473, A
396	5	1.2	20	10	US-09-931-700-7	Sequence 7, Appl	469	5	1.2	41	10	US-09-864-761-35881	Sequence 35881, A
397	5	1.2	21	9	US-10-042-431-23	Sequence 23, Appl	470	5	1.2	41	10	US-09-864-761-37983	Sequence 37983, A
398	5	1.2	22	9	US-09-843-676-134	Sequence 134, App	471	5	1.2	42	9	US-09-912-628-11	Sequence 11, Appl
399	5	1.2	22	9	US-09-766-253-154	Sequence 154, App	472	5	1.2	42	10	US-09-864-761-33906	Sequence 33906, A
400	5	1.2	23	9	US-09-964-012-5	Sequence 5, Appl	473	5	1.2	42	10	US-09-864-761-34838	Sequence 34838, A
401	5	1.2	23	9	US-09-331-631A-26	Sequence 26, Appl	474	5	1.2	42	10	US-09-864-761-37105	Sequence 37105, A
402	5	1.2	23	10	US-09-864-761-44616	Sequence 44616, A	475	5	1.2	43	10	US-09-843-845-5	Sequence 5, Appl
403	5	1.2	23	10	US-09-864-761-48488	Sequence 48488, A	476	5	1.2	43	10	US-09-864-761-34329	Sequence 34329, A
404	5	1.2	24	10	US-09-864-761-42885	Sequence 42885, A	477	5	1.2	43	10	US-09-864-761-38668	Sequence 38668, A
405	5	1.2	24	10	US-09-864-761-45452	Sequence 45452, A	478	5	1.2	43	10	US-09-864-761-38668	Sequence 38668, A
406	5	1.2	24	10	US-09-864-761-47525	Sequence 47525, A	479	5	1.2	43	10	US-09-847-185-48	Sequence 48, Appl
407	5	1.2	24	10	US-09-871-974-8	Sequence 8, Appl	480	5	1.2	43	10	US-09-981-289-5	Sequence 5, Appl
408	5	1.2	25	9	US-10-079-478-23	Sequence 23, Appl	481	5	1.2	43	10	US-09-764-877-1705	Sequence 1705, App
409	5	1.2	25	9	US-10-079-478-26	Sequence 26, Appl	482	5	1.2	44	10	US-09-843-845-26	Sequence 26, Appl
410	5	1.2	25	10	US-09-821-984-26	Sequence 26, Appl	483	5	1.2	44	10	US-09-864-761-42846	Sequence 42846, A
411	5	1.2	25	10	US-09-864-761-40473	Sequence 40473, A	484	5	1.2	44	10	US-09-864-761-47795	Sequence 47795, A
412	5	1.2	25	10	US-09-938-700-4	Sequence 4, Appl	485	5	1.2	45	9	US-09-984-245-157	Sequence 157, App
413	5	1.2	26	10	US-09-848-164-97	Sequence 97, Appl	486	5	1.2	45	10	US-09-864-761-37190	Sequence 37190, A
414	5	1.2	26	10	US-09-864-761-45669	Sequence 45669, A	487	5	1.2	45	10	US-09-864-761-38381	Sequence 38381, A
415	5	1.2	26	10	US-09-916-230-21	Sequence 21, Appl	488	5	1.2	45	10	US-09-864-761-41017	Sequence 41017, A
416	5	1.2	27	10	US-09-886-404-11	Sequence 11, Appl	489	5	1.2	45	10	US-09-925-299-858	Sequence 858, App
417	5	1.2	27	10	US-09-864-761-36931	Sequence 36931, A	490	5	1.2	45	10	US-09-789-561-121	Sequence 121, App
418	5	1.2	27	10	US-09-864-761-38622	Sequence 38622, A	491	5	1.2	45	10	US-09-948-080-5	Sequence 5, Appl
419	5	1.2	28	10	US-09-864-761-35251	Sequence 35251, A	492	5	1.2	45	10	US-09-948-080-26	Sequence 26, Appl
420	5	1.2	28	10	US-09-864-761-41150	Sequence 41150, A	493	5	1.2	46	9	US-10-001-876-148	Sequence 148, App
421	5	1.2	28	10	US-09-864-761-42119	Sequence 42119, A	494	5	1.2	46	10	US-09-864-761-39606	Sequence 39606, A
422	5	1.2	28	10	US-09-929-818-38	Sequence 38, Appl	495	5	1.2	46	10	US-09-864-761-41528	Sequence 41528, A
423	5	1.2	29	8	US-08-873-601-30	Sequence 30, Appl	496	5	1.2	46	10	US-09-864-761-42309	Sequence 42309, A
424	5	1.2	29	9	US-09-792-630-83	Sequence 83, Appl	497	5	1.2	46	10	US-09-864-761-42328	Sequence 42328, A
425	5	1.2	29	9	US-10-080-376-83	Sequence 83, Appl	498	5	1.2	47	10	US-09-864-761-43495	Sequence 43495, A
426	5	1.2	29	9	US-10-061-395-31	Sequence 31, Appl	499	5	1.2	47	10	US-09-864-761-46450	Sequence 46450, A
427	5	1.2	29	10	US-09-157-748-34	Sequence 34, Appl	500	5	1.2	48	10	US-09-864-761-34756	Sequence 34756, A
428	5	1.2	29	10	US-09-904-380-24	Sequence 24, Appl	501	5	1.2	48	10	US-09-864-761-39977	Sequence 39977, A
429	5	1.2	29	10	US-09-916-940-34	Sequence 34, Appl	502	5	1.2	48	10	US-09-864-761-41609	Sequence 41609, A
430	5	1.2	30	8	US-08-969-137-6	Sequence 6, Appl	503	5	1.2	48	10	US-09-864-761-44133	Sequence 44133, A
431	5	1.2	32	10	US-09-864-761-46467	Sequence 46467, A	504	5	1.2	50	9	US-09-989-920-173	Sequence 173, App
432	5	1.2	32	10	US-09-737-379-7	Sequence 7, Appl	505	5	1.2	50	10	US-09-864-761-39537	Sequence 39537, A
433	5	1.2	33	10	US-09-864-761-39255	Sequence 39255, A	506	5	1.2	50	10	US-09-925-299-1382	Sequence 1382, App
434	5	1.2	33	10	US-09-864-761-45878	Sequence 45878, A	507	5	1.2	51	9	US-10-002-344A-158	Sequence 158, App
435	5	1.2	33	10	US-09-281-717-50	Sequence 50, Appl	508	5	1.2	51	10	US-09-864-761-39015	Sequence 39015, A
436	5	1.2	34	10	US-09-864-761-33843	Sequence 33843, A	509	5	1.2	51	10	US-09-864-761-45353	Sequence 45353, A
437	5	1.2	34	10	US-09-864-761-40591	Sequence 40591, A	510	5	1.2	51	10	US-09-864-761-48742	Sequence 48742, A
438	5	1.2	34	10	US-09-864-761-45164	Sequence 45164, A	511	5	1.2	51	10	US-09-864-761-48898	Sequence 48898, A
439	5	1.2	34	10	US-09-864-761-46156	Sequence 46156, A	512	5	1.2	51	10	US-09-925-297-566	Sequence 566, App
440	5	1.2	34	10	US-09-864-761-47097	Sequence 47097, A	513	5	1.2	51	10	US-09-867-550-1582	Sequence 1582, App
441	5	1.2	34	10	US-09-864-761-48511	Sequence 48511, A	514	5	1.2	51	10	US-09-764-860-316	Sequence 316, App
442	5	1.2	35	10	US-09-900-530A-4	Sequence 4, Appl	515	5	1.2	51	10	US-09-972-484-7	Sequence 7, Appl
443	5	1.2	35	9	US-09-843-676-133	Sequence 133, App	516	5	1.2	52	9	US-09-984-245-133	Sequence 133, App
444	5	1.2	35	9	US-09-765-253-133	Sequence 133, App	517	5	1.2	52	9	US-10-001-876-165	Sequence 165, App
445	5	1.2	35	10	US-09-864-761-35212	Sequence 35212, A	518	5	1.2	52	10	US-09-864-761-38688	Sequence 38688, A
446	5	1.2	36	10	US-09-864-761-45650	Sequence 45650, A	519	5	1.2	52	10	US-09-864-761-49089	Sequence 49089, A
447	5	1.2	36	10	US-09-864-761-47972	Sequence 47972, A	520	5	1.2	53	9	US-10-001-835-206	Sequence 206, App
448	5	1.2	37	8	US-08-873-601-12	Sequence 12, Appl	521	5	1.2	53	9	US-09-984-245-128	Sequence 128, App
449	5	1.2	37	9	US-09-792-630-65	Sequence 65, Appl	522	5	1.2	53	10	US-09-864-761-35425	Sequence 35425, A
450	5	1.2	37	9	US-10-080-376-65	Sequence 65, Appl	523	5	1.2	53	10	US-09-867-550-160	Sequence 160, App
451	5	1.2	37	9	US-10-061-395-13	Sequence 13, Appl	524	5	1.2	54	10	US-09-876-478-14	Sequence 14, Appl
452	5	1.2	37	10	US-09-157-748-19	Sequence 19, Appl	525	5	1.2	54	10	US-09-864-761-36688	Sequence 36688, A
453	5	1.2	37	10	US-09-864-761-39176	Sequence 39176, A	526	5	1.2	54	10	US-09-864-761-39008	Sequence 39008, A
454	5	1.2	37	10	US-09-864-761-48952	Sequence 48952, A	527	5	1.2	54	10	US-09-864-761-45050	Sequence 45050, A
455	5	1.2	37	10	US-09-916-940-16	Sequence 16, Appl	528	5	1.2	55	10	US-09-864-761-34815	Sequence 34815, A
456	5	1.2	37	10	US-09-922-503-9	Sequence 9, Appl	529	5	1.2	55	10	US-09-864-761-38937	Sequence 38937, A
457	5	1.2	38	9	US-09-834-998A-5	Sequence 5, Appl	530	5	1.2	56	10	US-09-030-619-208	Sequence 208, App

531	5	1.2	56	10	US-09-864-761-33548	Sequence 33548, A
532	5	1.2	56	10	US-09-764-877-1828	Sequence 1828, Ap
533	5	1.2	57	10	US-09-864-761-35973	Sequence 35973, A
534	5	1.2	57	10	US-09-864-761-37634	Sequence 37634, A
535	5	1.2	57	10	US-09-864-761-44674	Sequence 44674, A
536	5	1.2	57	10	US-09-925-297-578	Sequence 578, App
537	5	1.2	58	9	US-10-046-961-15	Sequence 15, Appl
538	5	1.2	58	10	US-09-864-761-37367	Sequence 37367, A
539	5	1.2	58	10	US-09-864-761-37378	Sequence 37378, A
540	5	1.2	58	10	US-09-864-761-39848	Sequence 39848, A
541	5	1.2	58	10	US-09-970-108-2	Sequence 2, Appl
542	5	1.2	58	12	US-10-001-843-216	Sequence 216, App
543	5	1.2	59	10	US-09-864-761-40951	Sequence 40951, A
544	5	1.2	59	10	US-09-864-761-41455	Sequence 41455, A
545	5	1.2	59	10	US-09-864-761-44398	Sequence 44398, A
546	5	1.2	59	10	US-09-864-761-47493	Sequence 47493, A
547	5	1.2	59	10	US-09-948-080-14	Sequence 14, Appl
548	5	1.2	59	10	US-09-764-877-1638	Sequence 1638, Ap
549	5	1.2	59	10	US-09-925-300-1862	Sequence 1862, Ap
550	5	1.2	60	9	US-09-965-528-5	Sequence 5, Appl
551	5	1.2	60	10	US-09-864-761-48822	Sequence 48822, A
552	5	1.2	61	10	US-09-864-761-33346	Sequence 33346, A
553	5	1.2	61	10	US-09-864-761-36791	Sequence 36791, A
554	5	1.2	61	10	US-09-864-847-530	Sequence 530, App
555	5	1.2	62	10	US-09-864-761-37727	Sequence 37727, A
556	5	1.2	62	10	US-09-863-693-24	Sequence 24, Appl
557	5	1.2	62	10	US-09-764-847-572	Sequence 572, Appl
558	5	1.2	63	10	US-09-864-761-33825	Sequence 33825, A
559	5	1.2	63	10	US-09-864-761-36425	Sequence 36425, A
560	5	1.2	63	10	US-09-864-761-40178	Sequence 40178, A
561	5	1.2	63	10	US-09-864-761-40629	Sequence 40629, A
562	5	1.2	63	10	US-09-864-761-40966	Sequence 40966, A
563	5	1.2	63	10	US-09-864-761-47058	Sequence 47058, A
564	5	1.2	64	8	US-08-424-550B-384	Sequence 384, App
565	5	1.2	64	10	US-09-864-761-33358	Sequence 33358, A
566	5	1.2	64	10	US-09-864-761-33590	Sequence 33590, A
567	5	1.2	64	10	US-09-864-761-37985	Sequence 37985, A
568	5	1.2	64	10	US-09-864-761-38188	Sequence 38188, A
569	5	1.2	64	10	US-09-864-761-39913	Sequence 39913, A
570	5	1.2	64	10	US-09-864-761-43388	Sequence 43388, A
571	5	1.2	65	10	US-09-864-761-35895	Sequence 35895, A
572	5	1.2	65	10	US-09-864-761-37817	Sequence 37817, A
573	5	1.2	65	10	US-09-864-761-38648	Sequence 38648, A
574	5	1.2	65	10	US-09-864-761-39273	Sequence 39273, A
575	5	1.2	65	10	US-09-864-761-48982	Sequence 48982, A
576	5	1.2	65	10	US-09-864-761-33621	Sequence 33621, A
577	5	1.2	66	10	US-09-864-761-36044	Sequence 36044, A
578	5	1.2	66	10	US-09-864-761-41822	Sequence 41822, A
579	5	1.2	66	10	US-09-864-761-46973	Sequence 46973, A
580	5	1.2	66	10	US-09-942-446-11	Sequence 11, Appl
581	5	1.2	67	10	US-09-843-845-17	Sequence 17, Appl
582	5	1.2	67	10	US-09-864-761-44418	Sequence 44418, A
583	5	1.2	67	10	US-09-864-761-45693	Sequence 45693, A
584	5	1.2	67	10	US-09-815-242-11336	Sequence 11336, A
585	5	1.2	67	10	US-09-815-242-11510	Sequence 11510, A
586	5	1.2	67	10	US-09-912-962-33	Sequence 33, Appl
587	5	1.2	68	10	US-09-864-761-34777	Sequence 34777, A
588	5	1.2	68	10	US-09-864-761-35623	Sequence 35623, A
589	5	1.2	68	10	US-09-864-761-41099	Sequence 41099, A
590	5	1.2	68	10	US-09-864-761-44728	Sequence 44728, A
591	5	1.2	68	10	US-09-864-761-46362	Sequence 46362, A
592	5	1.2	68	10	US-09-867-550-876	Sequence 876, App
593	5	1.2	68	10	US-09-764-877-1099	Sequence 1099, App
594	5	1.2	69	10	US-09-864-761-34136	Sequence 34136, A
595	5	1.2	69	10	US-09-864-761-38134	Sequence 38134, A
596	5	1.2	69	10	US-09-867-550-1122	Sequence 1122, App
597	5	1.2	70	10	US-09-864-761-37066	Sequence 37066, A
598	5	1.2	70	10	US-09-864-761-41178	Sequence 41178, A
599	5	1.2	70	10	US-09-864-761-43163	Sequence 43163, A
600	5	1.2	70	10	US-09-864-761-44434	Sequence 44434, A
601	5	1.2	70	10	US-09-925-299-993	Sequence 993, App
602	5	1.2	70	10	US-09-894-882-161	Sequence 161, App
603	5	1.2	71	9	US-09-984-245-219	Sequence 219, App
604	5	1.2	71	10	US-09-864-761-40253	Sequence 40253, A
605	5	1.2	71	10	US-09-894-882-74	Sequence 74, Appl
606	5	1.2	71	10	US-09-894-882-119	Sequence 119, App
607	5	1.2	71	10	US-09-808-212-114	Sequence 114, Appl
608	5	1.2	72	8	US-08-424-550B-233	Sequence 233, App
609	5	1.2	72	9	US-09-770-528-4	Sequence 4, Appl
610	5	1.2	72	10	US-09-864-761-33388	Sequence 33388, A
611	5	1.2	72	10	US-09-815-242-13220	Sequence 13220, A
612	5	1.2	72	10	US-09-867-550-1414	Sequence 1414, App
613	5	1.2	72	10	US-09-731-272-251	Sequence 251, App
614	5	1.2	72	10	US-09-894-882-158	Sequence 158, App
615	5	1.2	73	10	US-09-864-761-35171	Sequence 35171, A
616	5	1.2	73	10	US-09-764-869-824	Sequence 824, App
617	5	1.2	73	10	US-09-731-872-155	Sequence 422, App
618	5	1.2	73	10	US-09-894-882-1452	Sequence 1452, App
619	5	1.2	73	10	US-09-894-882-164	Sequence 164, App
620	5	1.2	73	10	US-09-894-882-167	Sequence 167, App
621	5	1.2	73	10	US-09-894-882-173	Sequence 173, App
622	5	1.2	73	10	US-09-894-882-176	Sequence 176, App
623	5	1.2	73	10	US-09-900-530A-27	Sequence 27, Appl
624	5	1.2	73	10	US-09-764-877-1692	Sequence 1692, App
625	5	1.2	74	10	US-09-864-761-34188	Sequence 34188, A
626	5	1.2	74	10	US-09-864-761-37524	Sequence 37524, A
627	5	1.2	74	10	US-09-864-761-44238	Sequence 44238, A
628	5	1.2	74	10	US-09-898-416-17	Sequence 16, Appl
629	5	1.2	74	10	US-09-808-212A-16	Sequence 1810, App
630	5	1.2	74	10	US-09-764-877-1810	Sequence 1810, App
631	5	1.2	75	10	US-09-864-761-46556	Sequence 46556, A
632	5	1.2	75	10	US-09-894-882-302	Sequence 302, App
633	5	1.2	75	10	US-09-894-882-305	Sequence 305, App
634	5	1.2	75	10	US-09-894-882-308	Sequence 308, App
635	5	1.2	75	10	US-09-894-882-311	Sequence 311, App
636	5	1.2	75	10	US-09-894-882-314	Sequence 314, App
637	5	1.2	75	10	US-09-894-882-317	Sequence 317, App
638	5	1.2	75	10	US-09-764-877-1734	Sequence 1734, App
639	5	1.2	76	10	US-09-729-835-58	Sequence 58, Appl
640	5	1.2	76	10	US-09-864-761-41556	Sequence 41556, A
641	5	1.2	76	10	US-09-925-297-469	Sequence 469, App
642	5	1.2	76	10	US-09-867-550-1562	Sequence 1562, App
643	5	1.2	76	10	US-09-939-880-284	Sequence 284, App
644	5	1.2	76	10	US-09-963-896-5	Sequence 5, Appl
645	5	1.2	76	10	US-09-972-884-10	Sequence 10, Appl
646	5	1.2	76	10	US-09-764-847-826	Sequence 826, App
647	5	1.2	76	10	US-09-764-847-826	Sequence 826, App
648	5	1.2	77	9	US-09-970-033-11	Sequence 11, Appl
649	5	1.2	77	10	US-09-016-250C-8	Sequence 8, Appl
650	5	1.2	77	10	US-09-864-761-39700	Sequence 39700, A
651	5	1.2	77	10	US-09-016-869B-8	Sequence 8, Appl
652	5	1.2	77	10	US-09-925-300-1120	Sequence 1120, App
653	5	1.2	78	10	US-09-916-230-24	Sequence 24, Appl
654	5	1.2	78	10	US-09-764-877-1800	Sequence 1800, App
655	5	1.2	79	9	US-10-012-886-563	Sequence 563, App
656	5	1.2	79	9	US-09-895-793-563	Sequence 563, App
657	5	1.2	79	9	US-09-895-793-563	Sequence 563, App
658	5	1.2	79	9	US-09-895-793-563	Sequence 563, App
659	5	1.2	79	10	US-09-759-143-563	Sequence 563, App
660	5	1.2	79	10	US-09-864-761-42101	Sequence 42101, A
661	5	1.2	79	10	US-09-780-669-563	Sequence 563, App
662	5	1.2	79	10	US-09-764-869-672	Sequence 672, App
663	5	1.2	79	10	US-09-764-869-889	Sequence 889, App
664	5	1.2	79	10	US-09-822-827-563	Sequence 563, App
665	5	1.2	79	10	US-09-867-550-718	Sequence 718, App
666	5	1.2	79	10	US-09-867-550-734	Sequence 734, App
667	5	1.2	79	10	US-09-867-550-734	Sequence 734, App
668	5	1.2	80	9	US-10-114-893-56	Sequence 56, Appl
669	5	1.2	80	10	US-09-864-761-34436	Sequence 34436, A
670	5	1.2	80	10	US-09-864-761-36156	Sequence 36156, A
671	5	1.2	80	10	US-09-864-761-36364	Sequence 36364, A
672	5	1.2	80	10	US-09-864-761-48387	Sequence 48387, A
673	5	1.2	80	10	US-09-864-761-48470	Sequence 48470, A
674	5	1.2	80	10	US-09-764-447-526	Sequence 447, App
675	5	1.2	81	10	US-09-925-300-1561	Sequence 1561, App
676	5	1.2	81	10	US-09-864-761-44564	Sequence 44564, A
677	5	1.2	82	10	US-09-864-761-42676	Sequence 42676, A

677	5	1.2	82	10	US-09-864-761-47712	Sequence 47712, A	750	1.2	100	9	US-09-894-030-3	Sequence 3, Appli
678	5	1.2	83	9	US-09-764-868-376	Sequence 976, App	751	1.2	100	10	US-09-893-600-3	Sequence 3, Appli
679	5	1.2	83	9	US-10-016-1574-223	Sequence 223, App	752	1.2	100	10	US-09-864-761-33882	Sequence 33882, A
680	5	1.2	83	10	US-09-864-761-36915	Sequence 36915, A	753	1.2	100	10	US-09-893-499-3	Sequence 3, Appli
681	5	1.2	83	10	US-09-864-761-47063	Sequence 47063, A	754	1.2	100	10	US-09-885-478-16	Sequence 16, Appl
682	5	1.2	83	10	US-09-864-761-48330	Sequence 48330, A	755	1.2	100	10	US-09-885-478-17	Sequence 17, Appl
683	5	1.2	83	10	US-09-864-761-48951	Sequence 48951, A	756	1.2	100	10	US-09-922-261-62	Sequence 62, Appl
684	5	1.2	83	10	US-09-764-853-649	Sequence 649, App	757	1.2	101	9	US-09-464-099A-17	Sequence 17, Appl
685	5	1.2	84	10	US-09-864-761-46733	Sequence 46733, A	758	1.2	101	10	US-09-861-696-17	Sequence 1619, Ap
686	5	1.2	84	10	US-09-925-299-837	Sequence 837, App	759	1.2	101	10	US-09-764-877-1619	Sequence 58, Appl
687	5	1.2	85	10	US-09-016-750C-11	Sequence 11, Appl	760	1.2	102	8	US-08-424-550B-58	Sequence 6, Appli
688	5	1.2	85	10	US-09-864-761-38616	Sequence 38616, A	761	1.2	102	9	US-10-004-381-6	Sequence 48570, A
689	5	1.2	85	10	US-09-864-761-45415	Sequence 45415, A	762	1.2	102	10	US-09-864-761-48570	Sequence 1719, Ap
690	5	1.2	85	10	US-09-764-869-1024	Sequence 1024, Ap	763	1.2	102	10	US-09-925-301-1719	Sequence 1435, Ap
691	5	1.2	85	10	US-09-016-869B-11	Sequence 11, Appl	764	1.2	103	10	US-09-925-301-1435	Sequence 5361, Ap
692	5	1.2	85	10	US-09-016-869B-34	Sequence 34, Appl	765	1.2	103	10	US-09-815-243-5361	Sequence 507, App
693	5	1.2	85	10	US-09-881-752A-32	Sequence 32, Appl	766	1.2	103	10	US-09-925-297-507	Sequence 4, Appli
694	5	1.2	86	9	US-09-922-199A-18	Sequence 18, Appl	767	1.2	104	9	US-09-865-528-4	Sequence 24, Appl
695	5	1.2	86	10	US-09-864-761-35396	Sequence 35396, A	768	1.2	104	10	US-09-835-788A-24	Sequence 432, App
696	5	1.2	86	10	US-09-764-870-442	Sequence 442, App	769	1.2	104	10	US-09-939-980-432	Sequence 111, App
697	5	1.2	87	10	US-09-925-300-1077	Sequence 1077, Ap	770	1.2	104	10	US-09-764-855-111	Sequence 13, Appl
698	5	1.2	88	9	US-10-016-1574-213	Sequence 213, App	771	1.2	105	9	US-09-464-099A-13	Sequence 13, Appl
699	5	1.2	88	10	US-09-864-761-34181	Sequence 34181, A	772	1.2	105	10	US-09-861-696-13	Sequence 13, Appl
700	5	1.2	88	10	US-09-016-869B-33	Sequence 33, Appl	773	1.2	105	10	US-09-766-396-26	Sequence 26, Appl
701	5	1.2	89	9	US-10-001-876-193	Sequence 193, App	774	1.2	105	10	US-09-955-807-2	Sequence 2, Appli
702	5	1.2	89	10	US-09-864-761-36646	Sequence 36646, A	775	1.2	105	10	US-09-955-807-14	Sequence 14, Appl
703	5	1.2	89	10	US-09-764-869-920	Sequence 920, App	776	1.2	105	12	US-10-062-375-26	Sequence 26, Appl
704	5	1.2	89	10	US-09-016-869B-32	Sequence 32, Appl	777	1.2	106	10	US-09-864-761-40293	Sequence 40293, A
705	5	1.2	90	10	US-09-864-761-35453	Sequence 35453, A	778	1.2	106	10	US-09-815-242-12639	Sequence 12639, A
706	5	1.2	90	10	US-09-864-761-43431	Sequence 43431, A	779	1.2	106	10	US-09-815-242-12881	Sequence 12881, A
707	5	1.2	90	10	US-09-864-761-46373	Sequence 46373, A	780	1.2	106	10	US-09-894-998-14	Sequence 14, Appl
708	5	1.2	90	10	US-09-731-872-375	Sequence 375, App	781	1.2	106	12	US-09-764-847-666	Sequence 666, App
709	5	1.2	90	12	US-10-071-751-25	Sequence 25, Appl	782	1.2	106	12	US-10-062-254-310	Sequence 310, App
710	5	1.2	91	10	US-09-864-761-33404	Sequence 33404, A	783	1.2	107	10	US-10-062-254-316	Sequence 316, App
711	5	1.2	91	10	US-09-864-761-38421	Sequence 38421, A	784	1.2	107	10	US-09-864-761-45996	Sequence 45996, A
712	5	1.2	93	10	US-09-815-242-11552	Sequence 11552, A	785	1.2	107	10	US-09-764-877-1551	Sequence 1551, Ap
713	5	1.2	93	10	US-09-942-446-13	Sequence 13, Appl	786	1.2	108	9	US-09-769-145-4	Sequence 4, Appli
714	5	1.2	93	10	US-09-867-550-804	Sequence 804, App	787	1.2	108	9	US-09-769-145-4	Sequence 5, Appli
715	5	1.2	94	10	US-09-864-761-33435	Sequence 33435, A	788	1.2	108	10	US-09-925-301-1415	Sequence 1415, Ap
716	5	1.2	94	10	US-09-864-761-40640	Sequence 40640, A	789	1.2	108	10	US-09-867-550-552	Sequence 552, App
717	5	1.2	94	10	US-09-864-761-44711	Sequence 44711, A	790	1.2	108	10	US-09-881-752A-236	Sequence 236, App
718	5	1.2	94	10	US-09-864-761-48196	Sequence 48196, A	791	1.2	108	12	US-10-006-982-4	Sequence 4, Appli
719	5	1.2	94	10	US-09-864-761-48196	Sequence 48196, A	792	1.2	108	12	US-10-006-982-5	Sequence 5, Appli
720	5	1.2	94	10	US-09-815-242-5747	Sequence 5747, Ap	793	1.2	109	9	US-09-764-868-1139	Sequence 1139, Ap
721	5	1.2	95	9	US-09-993-333-12	Sequence 12, Appl	794	1.2	109	10	US-09-796-766-16	Sequence 16, Appl
722	5	1.2	95	10	US-09-886-468-26	Sequence 26, Appl	795	1.2	109	10	US-09-822-263-10	Sequence 10, Appl
723	5	1.2	95	10	US-09-956-425-18	Sequence 18, Appl	796	1.2	109	10	US-09-822-263-10	Sequence 10, Appl
724	5	1.2	95	10	US-09-956-425-21	Sequence 21, Appl	797	1.2	109	10	US-09-764-847-985	Sequence 985, App
725	5	1.2	95	10	US-09-956-425-22	Sequence 22, Appl	798	1.2	110	9	US-09-950-933A-69	Sequence 69, Appl
726	5	1.2	95	10	US-09-925-301-875	Sequence 875, App	799	1.2	110	10	US-09-925-301-1563	Sequence 1563, App
727	5	1.2	95	10	US-09-764-877-1610	Sequence 1610, Ap	800	1.2	111	10	US-09-764-853-454	Sequence 454, App
728	5	1.2	96	9	US-09-981-876-195	Sequence 195, App	801	1.2	112	10	US-09-216-393-83	Sequence 83, Appl
729	5	1.2	96	9	US-09-981-876-249	Sequence 249, App	802	1.2	112	10	US-09-919-473-10	Sequence 10, Appl
730	5	1.2	96	10	US-09-864-761-33793	Sequence 33793, A	803	1.2	113	9	US-09-263-959-328	Sequence 328, App
731	5	1.2	96	10	US-09-746-801A-41	Sequence 41, Appl	804	1.2	113	10	US-09-851-138-90	Sequence 90, Appl
732	5	1.2	96	10	US-09-764-847-661	Sequence 661, App	805	1.2	113	10	US-09-864-761-33818	Sequence 33818, A
733	5	1.2	97	9	US-10-036-041-16	Sequence 16, Appl	806	1.2	113	10	US-09-916-790-27	Sequence 27, Appl
734	5	1.2	97	10	US-09-925-299-1407	Sequence 1407, Ap	807	1.2	113	10	US-09-925-300-1363	Sequence 1363, Ap
735	5	1.2	97	12	US-10-036-342-16	Sequence 16, Appl	808	1.2	114	10	US-09-864-761-34736	Sequence 34736, A
736	5	1.2	97	12	US-10-052-586-468	Sequence 468, App	809	1.2	114	10	US-09-925-299-976	Sequence 976, App
737	5	1.2	98	9	US-09-970-033-2	Sequence 2, Appli	810	1.2	114	10	US-09-925-297-669	Sequence 669, App
738	5	1.2	98	10	US-09-205-658-58	Sequence 58, Appl	811	1.2	114	10	US-09-893-737-96	Sequence 96, Appl
739	5	1.2	98	10	US-09-844-353A-58	Sequence 58, Appl	812	1.2	114	10	US-09-764-847-505	Sequence 505, App
740	5	1.2	98	10	US-09-864-761-37789	Sequence 37789, A	813	1.2	114	10	US-09-263-959-327	Sequence 327, App
741	5	1.2	98	10	US-09-864-761-41798	Sequence 41798, A	814	1.2	115	9	US-09-987-675-4	Sequence 4, Appli
742	5	1.2	98	10	US-09-864-761-47485	Sequence 47485, A	815	1.2	115	10	US-09-864-761-47275	Sequence 1575, Ap
743	5	1.2	98	10	US-09-764-846-229	Sequence 229, App	816	1.2	115	10	US-09-925-301-1575	Sequence 58, Appl
744	5	1.2	98	10	US-09-772-719-50	Sequence 50, Appl	817	1.2	115	10	US-09-922-261-58	Sequence 4, Appli
745	5	1.2	98	10	US-09-263-959-1190	Sequence 1190, Ap	818	1.2	115	10	US-09-987-655-4	Sequence 2, Appli
746	5	1.2	99	9	US-09-942-087A-15	Sequence 15, Appl	819	1.2	116	9	US-10-029-654-2	Sequence 362, App
747	5	1.2	99	10	US-09-779-233-3	Sequence 3, Appli	820	1.2	116	10	US-10-079-623-362	Sequence 362, App
748	5	1.2	99	10	US-09-844-508-11	Sequence 11, Appl	821	1.2	116	10	US-09-864-761-38064	Sequence 38064, A
749	5	1.2	99	10	US-09-925-300-1750	Sequence 1750, Ap	822	1.2	116	10	US-09-864-761-44632	Sequence 44632, A

823	5	1.2	117	9	US-10-000-639-5	Sequence 5, Appl1	896	5	1.2	131	10	US-09-895-728-6	Sequence 6, Appl1
824	5	1.2	117	10	US-09-864-761-37128	Sequence 37128, A	897	5	1.2	131	10	US-09-925-1000-1091	Sequence 1091, Ap
825	5	1.2	117	10	US-09-864-761-44197	Sequence 44197, A	898	5	1.2	132	9	US-09-764-866-941	Sequence 941, App
826	5	1.2	117	10	US-09-966-264-30	Sequence 30, Appl	899	5	1.2	132	10	US-09-764-869-1158	Sequence 1158, Ap
827	5	1.2	118	10	US-09-741-666-463	Sequence 463, App	900	5	1.2	132	10	US-09-731-821-597	Sequence 297, Appl
828	5	1.2	118	10	US-09-815-242-10337	Sequence 10337, A	901	5	1.2	132	10	US-09-922-261-54	Sequence 54, Appl
829	5	1.2	118	10	US-09-815-242-11097	Sequence 11097, A	902	5	1.2	133	10	US-09-764-877-1498	Sequence 1498, Ap
830	5	1.2	118	10	US-09-972-086-10	Sequence 10, Appl	903	5	1.2	133	10	US-09-864-761-43793	Sequence 43793, A
831	5	1.2	119	9	US-09-860-670-121	Sequence 121, Appl	904	5	1.2	133	10	US-09-925-297-646	Sequence 642, App
832	5	1.2	119	10	US-09-794-987-4	Sequence 4, Appl1	905	5	1.2	133	10	US-09-867-550-342	Sequence 346, App
833	5	1.2	119	10	US-09-813-459-15	Sequence 15, Appl	906	5	1.2	133	10	US-09-867-550-598	Sequence 598, App
834	5	1.2	120	9	US-09-881-752A-42	Sequence 42, Appl	907	5	1.2	133	10	US-09-746-491-10	Sequence 10, Appl
835	5	1.2	120	9	US-09-898-751A-14	Sequence 14, Appl	908	5	1.2	134	8	US-08-424-550B-189	Sequence 189, App
836	5	1.2	120	9	US-09-792-793A-87	Sequence 87, Appl	909	5	1.2	134	10	US-09-764-847-670	Sequence 670, App
837	5	1.2	120	10	US-09-864-761-35860	Sequence 35860, A	910	5	1.2	135	9	US-10-001-873-23	Sequence 33, Appl
838	5	1.2	120	10	US-09-939-980-357	Sequence 357, App	911	5	1.2	136	9	US-09-860-670-95	Sequence 95, Appl
839	5	1.2	120	10	US-09-931-381A-12	Sequence 12, Appl1	912	5	1.2	136	10	US-09-864-761-83352	Sequence 38352, A
840	5	1.2	121	9	US-09-797-941A-4	Sequence 4, Appl1	913	5	1.2	136	10	US-09-815-242-11019	Sequence 11019, A
841	5	1.2	121	9	US-09-965-528-22	Sequence 22, Appl	914	5	1.2	137	10	US-09-972-086-18	Sequence 18, Appl
842	5	1.2	121	10	US-09-864-761-34580	Sequence 34580, A	915	5	1.2	137	10	US-09-739-254-163	Sequence 163, App
843	5	1.2	121	10	US-09-864-761-43304	Sequence 43304, A	916	5	1.2	137	10	US-09-016-750C-4	Sequence 4, Appl1
844	5	1.2	121	10	US-09-925-301-1576	Sequence 1576, Ap	917	5	1.2	137	10	US-09-904-615-163	Sequence 163, App
845	5	1.2	121	10	US-09-925-299-1441	Sequence 1441, Ap	918	5	1.2	137	10	US-09-815-242-13204	Sequence 13204, A
846	5	1.2	122	10	US-09-163-88	Sequence 88, Appl	919	5	1.2	137	10	US-09-815-242-13528	Sequence 13528, A
847	5	1.2	122	10	US-09-864-761-48526	Sequence 48526, A	920	5	1.2	137	10	US-09-016-869B-4	Sequence 4, Appl1
848	5	1.2	122	10	US-09-811-284-148	Sequence 148, App	921	5	1.2	137	10	US-09-925-100-1786	Sequence 1786, Ap
849	5	1.2	123	10	US-09-864-761-38371	Sequence 38371, A	922	5	1.2	137	12	US-10-071-751-23	Sequence 23, Appl
850	5	1.2	123	10	US-09-864-761-42017	Sequence 42017, A	923	5	1.2	138	10	US-09-016-750C-13	Sequence 13, Appl
851	5	1.2	123	10	US-09-764-869-751	Sequence 751, App	924	5	1.2	138	10	US-09-867-550-2068	Sequence 2068, Ap
852	5	1.2	123	10	US-09-925-300-1501	Sequence 1501, Ap	925	5	1.2	138	10	US-09-016-869B-13	Sequence 13, Appl
853	5	1.2	124	9	US-09-860-670-100	Sequence 100, App	926	5	1.2	138	10	US-09-864-866-62	Sequence 42, Appl
854	5	1.2	124	10	US-09-925-302-460	Sequence 460, App	927	5	1.2	138	10	US-09-764-877-1934	Sequence 1934, Ap
855	5	1.2	124	10	US-09-864-761-44327	Sequence 44327, A	928	5	1.2	139	9	US-09-819-994-4	Sequence 4, Appl1
856	5	1.2	124	10	US-09-764-869-1180	Sequence 1180, Ap	929	5	1.2	139	10	US-09-864-761-34808	Sequence 34808, A
857	5	1.2	124	10	US-09-915-242-5423	Sequence 5423, Ap	930	5	1.2	139	10	US-09-922-261-52	Sequence 52, Appl
858	5	1.2	124	10	US-09-972-484-25	Sequence 25, Appl	931	5	1.2	139	10	US-09-850-165-17	Sequence 17, Appl
859	5	1.2	125	9	US-09-978-295A-454	Sequence 454, Appl	932	5	1.2	140	9	US-09-922-199A-11	Sequence 11, Appl
860	5	1.2	125	9	US-09-764-868-791	Sequence 791, App	933	5	1.2	140	10	US-09-007-093-4	Sequence 4, Appl1
861	5	1.2	125	9	US-09-978-697-454	Sequence 454, App	934	5	1.2	140	10	US-09-943-002-14	Sequence 14, Appl
862	5	1.2	125	9	US-09-978-192A-454	Sequence 454, App	935	5	1.2	140	10	US-09-864-761-3518	Sequence 3518, A
863	5	1.2	125	9	US-09-999-832A-454	Sequence 454, App	936	5	1.2	140	10	US-09-864-761-45411	Sequence 45411, A
864	5	1.2	125	10	US-09-016-750C-6	Sequence 6, Appl1	937	5	1.2	141	9	US-09-984-245-303	Sequence 303, App
865	5	1.2	125	10	US-09-925-302-849	Sequence 849, App	938	5	1.2	141	9	US-09-764-868-1035	Sequence 1035, Ap
866	5	1.2	125	10	US-09-854-105-2	Sequence 2, Appl1	939	5	1.2	141	9	US-10-013-379-10	Sequence 10, Appl
867	5	1.2	125	10	US-09-016-869B-6	Sequence 6, Appl1	940	5	1.2	141	10	US-09-073-009-15	Sequence 15, Appl
868	5	1.2	125	10	US-09-731-872-348	Sequence 348, App	941	5	1.2	141	10	US-09-864-761-35711	Sequence 35711, A
869	5	1.2	125	10	US-09-809-739-1	Sequence 1, Appl1	942	5	1.2	141	10	US-09-864-761-37903	Sequence 37903, A
870	5	1.2	125	10	US-09-925-300-1854	Sequence 1854, Ap	943	5	1.2	141	10	US-09-023-588-15	Sequence 15, Appl
871	5	1.2	126	9	US-10-029-180-88	Sequence 88, Appl	944	5	1.2	141	10	US-09-867-550-880	Sequence 880, App
872	5	1.2	126	9	US-10-014-717-12	Sequence 12, Appl	945	5	1.2	141	10	US-09-972-086-9	Sequence 9, Appl1
873	5	1.2	127	10	US-10-001-857-167	Sequence 167, App	946	5	1.2	141	10	US-09-793-306-15	Sequence 15, Appl
874	5	1.2	127	10	US-09-016-750C-14	Sequence 14, Appl	947	5	1.2	141	10	US-09-746-284-73	Sequence 23, Appl
875	5	1.2	127	10	US-09-815-242-10860	Sequence 10860, A	948	5	1.2	142	10	US-09-925-302-790	Sequence 790, App
876	5	1.2	127	10	US-09-016-869B-14	Sequence 14, Appl	949	5	1.2	142	10	US-09-815-242-10434	Sequence 10434, A
877	5	1.2	127	10	US-09-925-300-1853	Sequence 1853, Ap	950	5	1.2	142	10	US-09-815-242-11039	Sequence 11039, A
878	5	1.2	128	10	US-09-864-761-45953	Sequence 45953, A	951	5	1.2	142	10	US-09-924-256A-20	Sequence 20, Appl
879	5	1.2	128	10	US-09-864-761-48538	Sequence 48538, A	952	5	1.2	142	12	US-10-062-254-366	Sequence 366, App
880	5	1.2	129	10	US-09-911-761-5	Sequence 5, Appl1	953	5	1.2	143	10	US-09-864-761-40061	Sequence 40061, A
881	5	1.2	129	10	US-09-864-761-42770	Sequence 42770, A	954	5	1.2	143	10	US-09-925-299-1335	Sequence 1335, Ap
882	5	1.2	129	10	US-09-815-242-10521	Sequence 10521, A	955	5	1.2	143	10	US-09-764-877-1731	Sequence 1731, Ap
883	5	1.2	129	10	US-09-972-086-17	Sequence 17, Appl	956	5	1.2	144	9	US-10-012-896-480	Sequence 480, App
884	5	1.2	129	10	US-09-764-864-1090	Sequence 1090, Ap	957	5	1.2	144	9	US-09-895-793-480	Sequence 480, App
885	5	1.2	129	10	US-09-764-864-1511	Sequence 1511, Ap	958	5	1.2	144	9	US-09-895-814-480	Sequence 480, App
886	5	1.2	130	10	US-09-802-077-5	Sequence 5, Appl1	959	5	1.2	144	10	US-09-759-143-480	Sequence 480, App
887	5	1.2	130	10	US-09-802-096-5	Sequence 5, Appl1	960	5	1.2	144	10	US-09-780-669-480	Sequence 480, App
888	5	1.2	130	10	US-09-164-615-22	Sequence 22, Appl	961	5	1.2	144	10	US-09-740-668A-30	Sequence 30, Appl
889	5	1.2	131	8	US-09-737-379-6	Sequence 6, Appl1	962	5	1.2	144	10	US-09-822-827-480	Sequence 480, App
890	5	1.2	131	10	US-09-812-393A-44	Sequence 44, Appl1	963	5	1.2	145	9	US-09-764-904-48	Sequence 48, Appl
891	5	1.2	131	10	US-09-815-242-11056	Sequence 11056, A	964	5	1.2	145	10	US-09-864-761-33938	Sequence 33938, A
892	5	1.2	131	10	US-09-164-615-26	Sequence 26, Appl	965	5	1.2	145	10	US-09-764-860-562	Sequence 562, App
893	5	1.2	131	10	US-09-164-615-46	Sequence 46, Appl	966	5	1.2	146	10	US-09-925-301-846	Sequence 846, App
894	5	1.2	131	10	US-09-737-379-2	Sequence 2, Appl1	967	5	1.2	146	10	US-09-764-864-1147	Sequence 1147, Ap
895	5	1.2	131	10	US-09-908-322-60	Sequence 60, Appl	968	5	1.2	147	9	US-09-819-994-2	Sequence 2, Appl1

Sequence 22, Appl  
Sequence 33801, A  
Sequence 36066, A  
Sequence 430, App  
Sequence 2, Appli  
Sequence 14, Appl  
Sequence 972, App  
Sequence 35, Appl  
Sequence 1711, Ap  
Sequence 2, Appli  
Sequence 1056, Ap  
Sequence 11, Appl  
Sequence 3, Appli  
Sequence 20, Appl  
Sequence 25, Appl  
Sequence 384, App  
Sequence 13, Appl  
Sequence 188, App  
Sequence 9, Appli  
Sequence 122, App  
Sequence 2, Appli  
Sequence 4, Appli  
Sequence 10, Appl  
Sequence 77, Appl  
Sequence 77, Appl  
Sequence 856, App  
Sequence 22, Appl  
Sequence 58, Appl  
Sequence 528, App  
Sequence 778, App  
Sequence 11, Appl  
Sequence 152, App

1.2 147 10 US-09-216-393-22  
1.2 147 10 US-09-864-761-33801  
1.2 147 10 US-09-864-761-36066  
1.2 147 10 US-09-833-790-430  
1.2 147 10 US-09-978-199-2  
1.2 147 10 US-09-749-601A-14  
1.2 147 10 US-09-764-847-972  
1.2 148 10 US-09-016-869B-35  
1.2 148 10 US-09-764-877-1711  
1.2 149 10 US-09-272-162-2  
1.2 149 10 US-09-764-864-1056  
1.2 149 12 US-10-000-759A-11  
1.2 150 10 US-09-862-419-3  
1.2 150 10 US-09-834-795A-20  
1.2 150 10 US-09-799-777-25  
1.2 150 10 US-09-731-872-384  
1.2 150 10 US-09-931-381A-13  
1.2 150 12 US-10-001-879-188  
1.2 150 12 US-10-000-759A-9  
1.2 151 9 US-09-797-941A-2  
1.2 151 10 US-09-867-550-122  
1.2 152 12 US-10-094-214-4  
1.2 153 9 US-10-120-544A-10  
1.2 153 10 US-09-739-254-77  
1.2 153 10 US-09-904-615-77  
1.2 153 10 US-09-925-301-856  
1.2 153 10 US-09-905-176-22  
1.2 153 12 US-10-078-929-58  
1.2 154 10 US-09-939-980-528  
1.2 154 10 US-09-764-853-778  
1.2 154 10 US-09-775-046-11  
1.2 155 12 US-10-006-867-152

ALIGNMENTS

RESULT 1  
US-10-022-461-6  
; Sequence 6, Application US/10022461  
; Patent No. US20020142420A1  
; GENERAL INFORMATION:  
; APPLICANT: Haake, David A.  
; TITLE OF INVENTION: CLONED leptospira OUTER MEMBRANE PROTEIN  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Spensley Horn Jubas & Lubitz  
; STREET: 1880 Century Park East, Suite 500  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90067  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/022,461  
; FILING DATE: 30-Oct-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 09/175,229  
; FILING DATE: 1998-10-28  
; APPLICATION NUMBER: US 08/249,013  
; FILING DATE: 25-MAY-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Tumarkin Ph.D., Lisa A.,  
; REGISTRATION NUMBER: P-38,347  
; REFERENCE/DOCKET NUMBER: PD-3602  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 455-5100  
; TELEFAX: (619) 455-5110

; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 117 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; IMMEDIATE SOURCE:  
; CLONE: luta  
; FEATURE:  
; NAME/KEY: Protein  
; LOCATION: 1..117  
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-10-022-461-6  
  
Query Match 2.5%; Score 10; DB 12; Length 117;  
Best Local Similarity 100.0%; Pred. No. 0.029;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 154 TGGLINIVTK 163  
|||  
Db 60 TCGLINIVTK 69  
  
RESULT 2  
US-09-815-242-5085  
; Sequence 5085, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard  
; TITLE OF INVENTION: Identification of Essential Genes in  
; FILE REFERENCE: ELITRA.011A  
; CURRENT APPLICATION NUMBER: US/09/815,242  
; CURRENT FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; NUMBER OF SEQ ID NOS: 14110  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5085  
; LENGTH: 813  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-815-242-5085  
  
Query Match 2.2%; Score 9; DB 10; Length 813;  
Best Local Similarity 100.0%; Pred. No. 1.5;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 148 IYGSATGG 156  
|||  
Db 229 IYGSATGG 237



```
RESULT 3
US-09-815-242-5436
; Sequence 5436, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5436
; LENGTH: 285
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5436

Query Match      1.8%; Score 7; DB 10; Length 285;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      290 LNAIKGL 296
DB      243 LNAIKGL 249

RESULT 4
US-09-815-242-12145
; Sequence 12145, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12145
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12145

Query Match      1.8%; Score 7; DB 10; Length 293;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12145
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12145

Query Match      1.8%; Score 7; DB 10; Length 293;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      290 LNAIKGL 296
DB      246 LNAIKGL 252

RESULT 5
US-09-815-242-12799
; Sequence 12799, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12799
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12799

Query Match      1.8%; Score 7; DB 10; Length 293;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      290 LNAIKGL 296
DB      246 LNAIKGL 252
```

```
RESULT 6
; Sequence 17, Application US/09393634
; Patent No. US2002005197A1
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Adler, Jon Elliot
; APPLICANT: Ryba, Nick
; APPLICANT: Mueller, Ken
; APPLICANT: Hoon, Mark
; APPLICANT: The Regents of the University of California
; APPLICANT: The Government of the United States of America
; APPLICANT: as represented by the Secretary of the
; APPLICANT: Department of Health and Human Services
; TITLE OF INVENTION: SF, a No. US2002005197A1el Family of Taste Receptors
; FILE REFERENCE: 02307E-098000US
; CURRENT APPLICATION NUMBER: US/09/393,634
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Mus sp.
; FEATURE:
; OTHER INFORMATION: mouse GR04
US-09-393-634-17

Query Match 1.8%; Score 7; DB 10; Length 300;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 361 TLDAYTK 367
Db 63 TLDAYTK 69

RESULT 7
US-10-008-118A-8
; Sequence 8, Application US/10008118A
; Publication No. US20020187539A1
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Weng, Zude
; TITLE OF INVENTION: Plant MYB-Related Transcription Factors
; FILE REFERENCE: BB1280 USDIV
; CURRENT APPLICATION NUMBER: US/10/008,118A
; CURRENT FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/109,294
; PRIOR FILING DATE: 1998-11-20
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Glycine max
US-10-008-118A-8

Query Match 1.8%; Score 7; DB 9; Length 305;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 362 LDAYTKA 368
Db 258 LDAYTKA 264

RESULT 8
US-09-443-704-8
; Sequence 8, Application US/09443704
; Patent No. US20020066120A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: P2040P1
; CURRENT APPLICATION NUMBER: US/09/726,643
; CURRENT FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/137,725
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 72
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-042-141-72

Query Match 1.8%; Score 7; DB 9; Length 322;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 199 SENG NVL 205
Db 59 SENG NVL 65

RESULT 10
US-09-726-643-72
; Sequence 72, Application US/09726643
; Patent No. US2002028449A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: P2040P1
; CURRENT APPLICATION NUMBER: US/09/726,643
; CURRENT FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
```

```
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Liu, Zhan-Bin
; APPLICANT: Odell, Joan
; APPLICANT: Rafalski, Antoni
; APPLICANT: Shi, June
; APPLICANT: Weng, Zude
; TITLE OF INVENTION: Plant MYB-Related Transcription Factors
; FILE REFERENCE: BB1280 US NA
; CURRENT APPLICATION NUMBER: US/09/443,704
; CURRENT FILING DATE: 1999-11-19
; EARLIER APPLICATION NUMBER: 60/109,294
; EARLIER FILING DATE: No. US20020066120A1ember 20, 1998
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 8
; LENGTH: 305
; TYPE: PRT
; ORGANISM: Glycine max
US-09-443-704-8

Query Match 1.8%; Score 7; DB 10; Length 305;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 362 LDAYTKA 368
Db 258 LDAYTKA 264

RESULT 9
US-10-042-141-72
; Sequence 72, Application US/10042141
; Publication No. US20020183503A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: P2040P1
; CURRENT APPLICATION NUMBER: US/10/042,141
; CURRENT FILING DATE: 2002-01-11
; PRIOR APPLICATION NUMBER: 09/726,643
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: 60/137,725
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 72
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-042-141-72

Query Match 1.8%; Score 7; DB 9; Length 322;
Best Local Similarity 100.0%; Pred. No. 51;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 199 SENG NVL 205
Db 59 SENG NVL 65

RESULT 10
US-09-726-643-72
; Sequence 72, Application US/09726643
; Patent No. US2002028449A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 26 Human secreted proteins
; FILE REFERENCE: P2040P1
; CURRENT APPLICATION NUMBER: US/09/726,643
; CURRENT FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/15187
```

;; PRIOR FILING DATE: 2000-06-02  
;; PRIOR APPLICATION NUMBER: 60/137,725  
;; PRIOR FILING DATE: 1999-06-07  
;; NUMBER OF SEQ ID NOS: 190  
;; SOFTWARE: Patentln Ver. 2.0  
;; SEQ ID NO 72  
;; LENGTH: 322  
;; TYPE: PRF  
;; ORGANISM: Homo sapiens  
US-09-726-643-72

Query Match 1.8%; Score 7; DB 10; Length 322;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 SENGNTL 205  
DB 59 SENGNTL 65

RESULT 11  
US-09-828-478-4  
; Sequence 4, Application US/09828478  
; Patent No. US20020155528A1  
; GENERAL INFORMATION:

;; APPLICANT: Xiao, Yonghong  
;; TITLE OF INVENTION: Regulation of Human CysLT2-Like GPCR  
;; FILE REFERENCE: 04974.00458  
;; CURRENT APPLICATION NUMBER: US/09/828,478  
;; CURRENT FILING DATE: 2001-04-09

;; PRIOR APPLICATION NUMBER: 60/195,196  
;; PRIOR FILING DATE: 2000-04-07  
;; PRIOR APPLICATION NUMBER: 60/254,876  
;; PRIOR FILING DATE: 2000-12-13

;; NUMBER OF SEQ ID NOS: 16  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 4  
;; LENGTH: 339  
;; TYPE: PRF

;; ORGANISM: Homo sapiens  
US-09-828-478-4

Query Match 1.8%; Score 7; DB 9; Length 339;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 ALANRIT 55  
DB 276 ALANRIT 282

RESULT 12  
US-09-848-889-12

;; Sequence 12, Application US/09848889  
;; Patent No. US20020025555A1  
; GENERAL INFORMATION:

;; APPLICANT: Au-Young, Janice  
;; APPLICANT: Guegler, Karl J.  
;; APPLICANT: Cheng, Muzong

;; TITLE OF INVENTION: GPCR DIAGNOSTIC FOR BRAIN CANCER  
;; FILE REFERENCE: PC-0042 CIP  
;; CURRENT APPLICATION NUMBER: US/09/848,889  
;; CURRENT FILING DATE: 2001-05-03

;; NUMBER OF SEQ ID NOS: 12  
;; SOFTWARE: PERL Program  
;; SEQ ID NO 12  
;; LENGTH: 339  
;; TYPE: PRF

;; ORGANISM: Homo sapiens  
;; NAME/KEY: misc.feature  
; OTHER INFORMATION: Incyte ID No. US20020025555A1 9992700

US-09-848-889-12

Query Match 1.8%; Score 7; DB 10; Length 339;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 ALANRIT 55  
DB 276 ALANRIT 282

RESULT 13  
US-09-788-133-2

;; Sequence 2, Application US/09788133  
;; Patent No. US20020052001A1  
; GENERAL INFORMATION:

;; APPLICANT: GLAXO GROUP LTD  
;; TITLE OF INVENTION: ASSAY  
;; FILE REFERENCE: P79011

;; CURRENT APPLICATION NUMBER: US/09/788,133  
;; CURRENT FILING DATE: 2001-02-20  
;; NUMBER OF SEQ ID NOS: 2  
;; SOFTWARE: Patentln Ver. 2.1

;; SEQ ID NO 2  
;; LENGTH: 339  
;; TYPE: PRF  
;; ORGANISM: homo sapiens  
US-09-788-133-2

Query Match 1.8%; Score 7; DB 10; Length 339;  
Best Local Similarity 100.0%; Pred. No. 53;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 ALANRIT 55  
DB 276 ALANRIT 282

RESULT 14  
US-09-828-478-6

;; Sequence 6, Application US/09828478  
;; Patent No. US20020155528A1  
; GENERAL INFORMATION:

;; APPLICANT: Xiao, Yonghong  
;; TITLE OF INVENTION: Regulation of Human CysLT2-Like GPCR  
;; FILE REFERENCE: 04974.00458  
;; CURRENT APPLICATION NUMBER: US/09/828,478

;; CURRENT FILING DATE: 2001-04-09  
;; PRIOR APPLICATION NUMBER: 60/195,196  
;; PRIOR FILING DATE: 2000-04-07

;; PRIOR APPLICATION NUMBER: 60/254,876  
;; PRIOR FILING DATE: 2000-12-13  
;; NUMBER OF SEQ ID NOS: 16  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 6  
;; LENGTH: 367  
;; TYPE: PRF

;; ORGANISM: Homo sapiens  
US-09-828-478-6

Query Match 1.8%; Score 7; DB 9; Length 367;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 ALANRIT 55  
DB 304 ALANRIT 310

RESULT 15  
US-09-881-752A-74  
; Sequence 74, Application US/09881752A

; Patent No. US20020115078A1  
; GENERAL INFORMATION:  
; APPLICANT: Kleanthous, Harold  
; APPLICANT: Al-Gatawi, Anal  
; APPLICANT: Miller, Charles  
; APPLICANT: Tomb, Jean-Francois  
; APPLICANT: Oomen, Raymond P.  
; TITLE OF INVENTION: Identification of Polynucleotides  
; TITLE OF INVENTION: Encoding No. US20020115078A1el Helicobacter Polypeptides in the  
; TITLE OF INVENTION: Genome  
; FILE REFERENCE: 06132/041002  
; CURRENT APPLICATION NUMBER: US/09/881,752A  
; PRIOR FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: US 08/833,457  
; PRIOR FILING DATE: 1997-04-01  
; NUMBER OF SEQ ID NOS: 370  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 74  
; LENGTH: 386  
; TYPE: PRT  
; ORGANISM: Helicobacter pylori  
US-09-881-752A-74

Query Match 1.8%; Score 7; DB 10; Length 386;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 294 KGLSLSE 300  
Db 34 KGLSLSE 40

RESULT 16  
US-10-042-141-46  
; Sequence 46, Application US/10042141  
; Publication No. US20020183503A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: 26 Human secreted proteins  
; FILE REFERENCE: P2040P1  
; CURRENT APPLICATION NUMBER: US/10/042,141  
; CURRENT FILING DATE: 2002-01-11  
; PRIOR APPLICATION NUMBER: 09/726,643  
; PRIOR FILING DATE: 2000-12-01  
; PRIOR APPLICATION NUMBER: PCT/US00/15187  
; PRIOR FILING DATE: 2000-06-02  
; PRIOR APPLICATION NUMBER: 60/137,725  
; PRIOR FILING DATE: 1999-06-07  
; NUMBER OF SEQ ID NOS: 190  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 46  
; LENGTH: 453  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-042-141-46

Query Match 1.8%; Score 7; DB 9; Length 453;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 199 SENG NVL 205  
Db 68 SENG NVL 74

RESULT 17  
US-09-726-643-46  
; Sequence 46, Application US/09726643  
; Patent No. US2002028449A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: 26 Human secreted proteins  
; FILE REFERENCE: P2040P1

; CURRENT APPLICATION NUMBER: US/09/726,643  
; CURRENT FILING DATE: 2000-12-01  
; PRIOR APPLICATION NUMBER: PCT/US00/15187  
; PRIOR FILING DATE: 2000-06-02  
; PRIOR APPLICATION NUMBER: 60/137,725  
; PRIOR FILING DATE: 1999-06-07  
; NUMBER OF SEQ ID NOS: 190  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 46  
; LENGTH: 453  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-726-643-46

Query Match 1.8%; Score 7; DB 10; Length 453;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 199 SENG NVL 205  
Db 68 SENG NVL 74

RESULT 18  
US-10-042-141-82  
; Sequence 82, Application US/10042141  
; Publication No. US20020183503A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: 26 Human secreted proteins  
; FILE REFERENCE: P2040P1  
; CURRENT APPLICATION NUMBER: US/10/042,141  
; CURRENT FILING DATE: 2002-01-11  
; PRIOR APPLICATION NUMBER: 09/726,643  
; PRIOR FILING DATE: 2000-12-01  
; PRIOR APPLICATION NUMBER: PCT/US00/15187  
; PRIOR FILING DATE: 2000-06-02  
; PRIOR APPLICATION NUMBER: 60/137,725  
; PRIOR FILING DATE: 1999-06-07  
; NUMBER OF SEQ ID NOS: 190  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 82  
; LENGTH: 489  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: SITE  
; LOCATION: (18)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-10-042-141-82

Query Match 1.8%; Score 7; DB 9; Length 489;  
Best Local Similarity 100.0%; Pred. No. 74;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 199 SENG NVL 205  
Db 104 SENG NVL 110

RESULT 19  
US-09-726-643-82  
; Sequence 82, Application US/09726643  
; Patent No. US2002028449A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: 26 Human secreted proteins  
; FILE REFERENCE: P2040P1  
; CURRENT APPLICATION NUMBER: US/09/726,643  
; CURRENT FILING DATE: 2000-12-01  
; PRIOR APPLICATION NUMBER: PCT/US00/15187  
; PRIOR FILING DATE: 2000-06-02  
; PRIOR APPLICATION NUMBER: 60/137,725

```
; PRIOR FILING DATE: 1999-06-07
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 82
; LENGTH: 489
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (18)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-726-643-82
```

```
Query Match      1.8%; Score 7; DB 10; Length 489;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 199 SENGNTVL 205
DB 104 SENGNTVL 110
```

```
RESULT 20
US-09-925-300-1151
; Sequence 1151, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Ruben,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1151
; LENGTH: 489
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (18)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1151
```

```
Query Match      1.8%; Score 7; DB 10; Length 489;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 199 SENGNTVL 205
DB 104 SENGNTVL 110
```

```
RESULT 21
US-09-741-669-330
; Sequence 330, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
```

```
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 330
; LENGTH: 648
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-741-669-330
```

```
Query Match      1.8%; Score 7; DB 10; Length 648;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 14 LLSVAVT 20
DB 264 LLSVAVT 270
```

```
RESULT 22
US-09-815-242-10089
; Sequence 10089, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-23
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-10-23
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-11-27
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2000-12-22
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10089
; LENGTH: 648
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10089
```

```
Query Match      1.8%; Score 7; DB 10; Length 648;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 14 LLSVAVT 20
DB 264 LLSVAVT 270
```

```
RESULT 23
US-09-915-789A-13
; Sequence 13, Application US/09915789A
; Patent No. US20020168762A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Lieping
; TITLE OF INVENTION: B7-H3 AND B7-H4, NOVEL IMMUNOREGULATORY
```

; TITLE OF INVENTION: MOLECULES  
; FILE REFERENCE: 07039-219001  
; CURRENT APPLICATION NUMBER: US/09/915,789A  
; CURRENT FILING DATE: 2002-06-04  
; PRIOR APPLICATION NUMBER: US 60/220,991  
; PRIOR FILING DATE: 2000-07-27  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 24  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-915-789A-13

Query Match 1.5%; Score 6; DB 9; Length 24;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 115 GVPLTG 120  
|||||  
Db 18 GVPLTG 23

## RESULT 24

US-10-016-634A-104  
; Sequence 104, Application US/10016634A  
; Publication No. US20020192666A1  
; GENERAL INFORMATION:

APPLICANT: Sun, Yongming  
APPLICANT: Recipon, Herve  
APPLICANT: Ghosh, Malavika  
APPLICANT: Liu, Chenghua

; TITLE OF INVENTION: Compositions and Methods Relating to Colon Specific Genes and Pro  
; FILE REFERENCE: DEX-0255  
; CURRENT APPLICATION NUMBER: US/10/016,634A

; CURRENT FILING DATE: 2001-10-31

; PRIOR APPLICATION NUMBER: US 60/244,258

; PRIOR FILING DATE: 2000-10-31

; NUMBER OF SEQ ID NOS: 176

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 104

; LENGTH: 26

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-016-634A-104

Query Match 1.5%; Score 6; DB 9; Length 26;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 355 MNLPSA 360  
|||||  
Db 5 MNLPSA 10

## RESULT 25

US-09-864-761-48223  
; Sequence 48223, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:

APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

; FILE REFERENCE: Acomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 48223  
; LENGTH: 37  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AB019441.1  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.42  
US-09-864-761-48223

Query Match 1.5%; Score 6; DB 10; Length 37;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 SLPILLS 16  
|||||  
Db 7 SLPILLS 12

## RESULT 26

US-09-864-761-47372  
; Sequence 47372, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:

APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR

; FILE REFERENCE: Acomica-X-1

; CURRENT APPLICATION NUMBER: US/09/864,761

; CURRENT FILING DATE: 2001-05-23

; PRIOR APPLICATION NUMBER: US 60/180,312

; PRIOR FILING DATE: 2000-02-04

; PRIOR APPLICATION NUMBER: US 60/207,456

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: US 09/632,366

; PRIOR FILING DATE: 2000-08-03

```

; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 47372
; LENGTH: 40
; TYPE: prt
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL161646.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.63
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.48
US-09-864-761-47372

Query Match          1.5%; Score 6; DB 10; Length 40;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 QVQGSV 195
      |||||
Db 27 QVQGSV 32

RESULT 27
US-09-864-761-33872
; Sequence 33872, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
```

```

; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33872
; LENGTH: 42
; TYPE: prt
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007600.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 6.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.5
US-09-864-761-33872

Query Match          1.5%; Score 6; DB 10; Length 42;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 294 KGLSLG 299
      |||||
Db 33 KGLSLG 38

RESULT 28
US-09-864-761-40375
; Sequence 40375, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
```

;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 40375  
;; LENGTH: 42  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AL158049.2  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.3  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.5  
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.9  
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.3  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.3  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2  
;; OTHER INFORMATION: SWISSPROT HIT: P45988, EVALUATE 4.80e+00

US-09-864-761-40375

Query Match 1.5%; Score 6; DB 10; Length 42;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 179 GSKLSS 184  
|||  
DB 20 GSKLSS 25

RESULT 29

US-09-995-494-94  
;; Sequence 94, Application US/09995494  
;; Patent No. US20020127578A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Salceda, Susana  
;; APPLICANT: Macina, Roberto  
;; APPLICANT: Recipon, Harve  
;; APPLICANT: Caffierkey, Robert  
;; APPLICANT: Ali, Shujath

;; APPLICANT: Sun, Yongming  
;; APPLICANT: Liu, Chenchua  
;; APPLICANT: Chen, Sei-yu  
;; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific Genes and  
;; FILE REFERENCE: DEX-0293  
;; CURRENT APPLICATION NUMBER: US/09/995,494  
;; CURRENT FILING DATE: 2001-11-27  
;; PRIOR APPLICATION NUMBER: 60/253,176  
;; PRIOR FILING DATE: 2000-11-27  
;; NUMBER OF SEQ ID NOS: 115  
;; SOFTWARE: PatentIn version 3.1  
;; SEQ ID NO 94  
;; LENGTH: 45  
;; TYPE: PRT  
;; ORGANISM: Homo sapien  
;; US-09-995-494-94

Query Match 1.5%; Score 6; DB 10; Length 45;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 SVAVTQ 21  
|||  
DB 23 SVAVTQ 28

RESULT 30

US-09-864-761-42399  
;; Sequence 42399, Application US/09864761  
;; Patent No. US20020048763A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Penn, Sharron G.  
;; APPLICANT: Rank, David R.  
;; APPLICANT: Hanzel, David K.  
;; APPLICANT: Chen, Wensheng  
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
;; FILE REFERENCE: Aecmica-X-1  
;; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
;; CURRENT APPLICATION NUMBER: US/09/864,761  
;; CURRENT FILING DATE: 2001-05-23  
;; PRIOR APPLICATION NUMBER: US 60/180,312  
;; PRIOR FILING DATE: 2000-02-04  
;; PRIOR APPLICATION NUMBER: US 60/207,456  
;; PRIOR FILING DATE: 2000-05-26  
;; PRIOR APPLICATION NUMBER: US 09/632,366  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: GB 24263.6  
;; PRIOR FILING DATE: 2000-10-04  
;; PRIOR APPLICATION NUMBER: US 60/236,359  
;; PRIOR FILING DATE: 2000-09-27  
;; PRIOR APPLICATION NUMBER: PCT/US01/00666  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00667  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00664  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00669  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00665  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00668  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00663  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00662  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408



```

; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO: 42399
; LENGTH: 47
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC002044.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.9
; OTHER INFORMATION: SWISSPROT HIT: P48935, EVALUE 2.80e+00
; OTHER INFORMATION: EST_HUMAN HIT: BE741526.1, EVALUE 1.30e+00
US-09-864-761-42399
```

```

Query Match          1.5%; Score 6; DB 10; Length 47;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 352 LOSMNL 357

Db 34 LOSMNL 39

```

RESULT 31
US-09-864-761-39963
; Sequence 39963, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO: 39963
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005924.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EST_HUMAN HIT: AM402306.1, EVALUE 2.00e-07
US-09-864-761-39963
```

```

Query Match          1.5%; Score 6; DB 10; Length 50;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 35 ELEPVV 40

Db 13 ELEPVV 18

```

RESULT 32
US-09-864-761-45308
; Sequence 45308, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
```

```
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 45308
; LENGTH: 50
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC023296.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.74
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.73
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.71
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.91
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.74
; OTHER INFORMATION: SWISSPROT HIT: Q24248, EVALUE 1.60e+00
US-09-864-761-45308
```

```
Query Match 1.5%; Score 6; DB 10; Length 50;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 121 SRDISR 126
Db 2 SRDISR 7
```

```
RESULT 33
US-09-864-761-34894
; Sequence 34894, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wenheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34894
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009955.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.1
; OTHER INFORMATION: EST HUMAN HIT: BE868202.1, EVALUE 3.00e-16
; OTHER INFORMATION: SWISSPROT HIT: O18917, EVALUE 1.00e-25
US-09-864-761-34894
```

```
Query Match 1.5%; Score 6; DB 10; Length 56;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 280 LAVLFG 285
Db 14 LAVLFG 19
```

```
RESULT 34
US-09-779-451-76
; Sequence 76, Application US/09779451
; Patent No. US20020094521A1
; GENERAL INFORMATION:
; APPLICANT: Wild, Carl T.
; APPLICANT: Allaway, Graham P.
; TITLE OF INVENTION: Assay for Detection of Viral Fusion Inhibitors
; FILE REFERENCE: 1900.0300003
; CURRENT APPLICATION NUMBER: US/09/779,451
; CURRENT FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: US 60/235,901
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/181,543
; PRIOR FILING DATE: 2000-02-10
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 76
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Human parainfluenza virus 3
US-09-779-451-76
```

```
Query Match 1.5%; Score 6; DB 10; Length 56;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy 163 KSDLEE 168  
 Db 27 KSDLEE 32

## RESULT 35

US-10-035-408-4  
 ; Sequence 4, Application US/10035408  
 ; Patent No. US20020123117A1  
 ; GENERAL INFORMATION:

APPLICANT: WALLACH, David

BOLDIN, Mark P.

VARFOLOMEEV, Eugene E.

PANCER, Zeev

NETT, Igor

CONCHAROV, Tanya M.

WEINWURZEL, Henry

TITLE OF INVENTION: MODULATORS OF REGULATORY PROTEINS

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.

STREET: 419 Seventh Street N.W., Ste. 300

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20004

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/035,408

FILING DATE: 04-Jan-2002

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/894,626

FILING DATE: 09-Dec-1997

APPLICATION NUMBER: IL 112,742

FILING DATE: 22-FEB-1995

APPLICATION NUMBER: IL 115,289

FILING DATE: 13-SEP-1995

APPLICATION NUMBER: PCT/US96/02326

FILING DATE: 15-FEB-1996

ATTORNEY/AGENT INFORMATION:

NAME: BROWDY, Roger L.

REGISTRATION NUMBER: 25,618

REFERENCE/DOCKET NUMBER: WALLACH=17

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 628-5197

TELEFAX: (202) 737-3528

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 56 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 4:

US-10-035-408-4

Query Match 1.5%; Score 6; DB 12; Length 56;

Best Local Similarity 100.0%; Pred. No. 93;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 359 SATLDA 364

Db 37 SATLDA 42

RESULT 36

US-09-764-847-525

; Sequence 525, Application US/09764847

; Patent No. US20020132767A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PC009

; CURRENT APPLICATION NUMBER: US/09/764,847

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - consult PALM or file wrapper

; NUMBER OF SEQ ID NOS: 2003

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 525

; LENGTH: 58

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-764-847-525

Query Match 1.5%; Score 6; DB 10; Length 58;

Best Local Similarity 100.0%; Pred. No. 96;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 38 PVTIT 43

Db 2 PVTIT 7

RESULT 37

US-09-867-550-1698

; Sequence 1698, Application US/09867550

; Patent No. US2002082206A1

; GENERAL INFORMATION:

; APPLICANT: Leach, Martin D.

; APPLICANT: Mehraban, Fuad.

; APPLICANT: Conley, Pamela

; APPLICANT: Law, Debbie

; APPLICANT: Topper, James

; TITLE OF INVENTION: No. US2002082206A1 Polynucleotides from Atherogenic Cells and

; FILE REFERENCE: 21402-013 (Cura-313)

; CURRENT APPLICATION NUMBER: US/09/867,550

; CURRENT FILING DATE: 2001-09-20

; PRIOR APPLICATION NUMBER: USSN 60/208,427

; PRIOR FILING DATE: 2000-05-30

; NUMBER OF SEQ ID NOS: 2125

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 1698

; LENGTH: 70

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-867-550-1698

Query Match 1.5%; Score 6; DB 10; Length 70;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 31 LPTVEL 36

Db 33 LPTVEL 38

RESULT 38

US-09-764-877-1603

; Sequence 1603, Application US/09764877

; Patent No. US20020147140A1

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PC005

; CURRENT APPLICATION NUMBER: US/09/764,877

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - refer to PALM or file wrapper

; NUMBER OF SEQ ID NOS: 4031

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 1603

```
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (47)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (70)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-877-1603

Query Match          1.5%; Score 6; DB 10; Length 84;
Best Local Similarity 100.0%; Pred.No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 11 LSLPLLS 16
Db 31 LSLPLLS 36

RESULT 39
US-09-864-761-34118
; Sequence 34118, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
```

```
; SEQ ID NO 34118
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000154.1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EST HUMAN HIT: AUI19258.1, EVALUE 5.00e-27
; OTHER INFORMATION: SWISSPROT HIT: Q63470, EVALUE 5.00e-28
US-09-864-761-34118
```

```
Query Match          1.5%; Score 6; DB 10; Length 86;
Best Local Similarity 100.0%; Pred.No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 93 SSGTTS 98
Db 75 SSGTTS 80
```

## RESULT 40

```
US-09-982-405-2
; Sequence 2, Application US/09982405
; Patent No. US20020164764A1
; GENERAL INFORMATION:
; APPLICANT: Paul O. Sheppard
; APPLICANT: Betty A. Haldeman
; APPLICANT: Richard D. Holly
; TITLE OF INVENTION: Transmembrane Polypeptide Expressed by
; FILE REFERENCE: 98-43C1
; CURRENT APPLICATION NUMBER: US/09/982,405
; FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 09/631,073
; PRIOR FILING DATE: 2000-08-02
; PRIOR APPLICATION NUMBER: 09/394,767
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 60/100,865
; PRIOR FILING DATE: 1998-09-17
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-982-405-2
```

```
Query Match          1.5%; Score 6; DB 9; Length 92;
Best Local Similarity 100.0%; Pred.No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 10 LSLPLL 15
Db 45 LSLPLL 50
```

## RESULT 41

```
US-10-001-054-20
; Sequence 20, Application US/10001054
; Publication No. US20020192209A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Baker, Kevin
; APPLICANT: Goddard, Audrey
```

APPLICANT: Gurney, Austin  
APPLICANT: Hebert, Carolyn  
APPLICANT: Henzel, William  
APPLICANT: Kabakoff, Rhona  
APPLICANT: Shelton, David  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin  
APPLICANT: Wood, William  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITTING NEOPLASTIC  
FILE REFERENCE: P3034R1PCT  
CURRENT FILING DATE: 2001-11-30  
PRIOR APPLICATION NUMBER: 60/059114  
PRIOR FILING DATE: 1997-09-17  
PRIOR APPLICATION NUMBER: 60/079689  
PRIOR FILING DATE: 1998-03-27  
PRIOR APPLICATION NUMBER: 60/079920  
PRIOR FILING DATE: 1998-03-30  
PRIOR APPLICATION NUMBER: 60/082999  
PRIOR FILING DATE: 1998-04-24  
PRIOR APPLICATION NUMBER: 60/083545  
PRIOR FILING DATE: 1998-04-29  
PRIOR APPLICATION NUMBER: 60/085149  
PRIOR FILING DATE: 1998-05-12  
PRIOR APPLICATION NUMBER: 60/087607  
PRIOR FILING DATE: 1998-06-02  
PRIOR APPLICATION NUMBER: 60/088588  
PRIOR FILING DATE: 1998-06-11  
PRIOR APPLICATION NUMBER: 60/090691  
PRIOR FILING DATE: 1998-06-25  
PRIOR APPLICATION NUMBER: 60/096891  
PRIOR FILING DATE: 1998-08-17  
PRIOR APPLICATION NUMBER: 60/096894  
PRIOR FILING DATE: 1998-08-17  
PRIOR APPLICATION NUMBER: 60/099803  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: 60/100263  
PRIOR FILING DATE: 1998-09-14  
PRIOR APPLICATION NUMBER: 60/100390  
PRIOR FILING DATE: 1998-09-15  
PRIOR APPLICATION NUMBER: 60/101476  
PRIOR FILING DATE: 1998-09-23  
PRIOR APPLICATION NUMBER: 60/107783  
PRIOR FILING DATE: 1998-11-10  
PRIOR APPLICATION NUMBER: 60/108849  
PRIOR FILING DATE: 1998-11-18  
PRIOR APPLICATION NUMBER: 60/112420  
PRIOR FILING DATE: 1998-12-15  
PRIOR APPLICATION NUMBER: 60/113296  
PRIOR FILING DATE: 1998-12-22  
PRIOR APPLICATION NUMBER: 60/115554  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/115558  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/116533  
PRIOR FILING DATE: 1999-01-20  
PRIOR APPLICATION NUMBER: 60/123618  
PRIOR FILING DATE: 1999-03-10  
PRIOR APPLICATION NUMBER: 60/131294  
PRIOR FILING DATE: 1999-04-07  
PRIOR APPLICATION NUMBER: 60/140650  
PRIOR FILING DATE: 1999-06-22  
PRIOR APPLICATION NUMBER: 60/141037  
PRIOR FILING DATE: 1999-06-23  
PRIOR APPLICATION NUMBER: 60/144758  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: 60/162506  
PRIOR FILING DATE: 1999-10-29  
PRIOR APPLICATION NUMBER: 60/170262  
PRIOR FILING DATE: 1999-12-09  
PRIOR APPLICATION NUMBER: 60/187202  
PRIOR FILING DATE: 2000-03-03

PRIOR APPLICATION NUMBER: 60/209832  
PRIOR FILING DATE: 2000-06-05  
PRIOR APPLICATION NUMBER: 60/232887  
PRIOR FILING DATE: 2000-09-15  
PRIOR APPLICATION NUMBER: 09/180997  
PRIOR FILING DATE: 1998-11-19  
PRIOR APPLICATION NUMBER: 09/218517  
PRIOR FILING DATE: 1998-12-22  
PRIOR APPLICATION NUMBER: 09/284291  
PRIOR FILING DATE: 1999-04-12  
PRIOR APPLICATION NUMBER: 09/380137  
PRIOR FILING DATE: 1999-08-25  
PRIOR APPLICATION NUMBER: 09/380138  
PRIOR FILING DATE: 1999-08-25  
PRIOR APPLICATION NUMBER: 09/380913  
PRIOR FILING DATE: 1999-09-09  
PRIOR APPLICATION NUMBER: 09/403297  
PRIOR FILING DATE: 1999-10-18  
PRIOR APPLICATION NUMBER: 09/423741  
PRIOR FILING DATE: 1999-11-10  
PRIOR APPLICATION NUMBER: 09/709238  
PRIOR FILING DATE: 2000-11-08  
PRIOR APPLICATION NUMBER: 09/802706  
PRIOR FILING DATE: 2001-03-09  
PRIOR APPLICATION NUMBER: 09/866034  
PRIOR FILING DATE: 2001-05-25  
PRIOR APPLICATION NUMBER: 09/872035  
PRIOR FILING DATE: 2001-06-01  
PRIOR APPLICATION NUMBER: 09/882636  
PRIOR FILING DATE: 2001-06-14  
PRIOR APPLICATION NUMBER: 09/918585  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: 09/924419  
PRIOR FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: 09/927796  
PRIOR FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: 09/929404  
PRIOR FILING DATE: 2001-08-13  
PRIOR APPLICATION NUMBER: 09/941992  
PRIOR FILING DATE: 2001-08-28  
PRIOR APPLICATION NUMBER: 09/946374  
PRIOR FILING DATE: 2001-09-04  
PRIOR APPLICATION NUMBER: PCT/US98/18824  
PRIOR FILING DATE: 1998-09-10  
PRIOR APPLICATION NUMBER: PCT/US99/00106  
PRIOR FILING DATE: 1999-01-05  
PRIOR APPLICATION NUMBER: PCT/US99/05028  
PRIOR FILING DATE: 1999-03-08  
PRIOR APPLICATION NUMBER: PCT/US99/08615  
PRIOR FILING DATE: 1999-04-20  
PRIOR APPLICATION NUMBER: PCT/US99/12252  
PRIOR FILING DATE: 1999-06-02  
PRIOR APPLICATION NUMBER: PCT/US99/20111  
PRIOR FILING DATE: 1999-09-01  
PRIOR APPLICATION NUMBER: PCT/US99/20594  
PRIOR FILING DATE: 1999-09-08  
PRIOR APPLICATION NUMBER: PCT/US99/28313  
PRIOR FILING DATE: 1999-11-30  
PRIOR APPLICATION NUMBER: PCT/US99/28551  
PRIOR FILING DATE: 1999-12-02  
PRIOR APPLICATION NUMBER: PCT/US99/28634  
PRIOR FILING DATE: 1999-12-01  
PRIOR APPLICATION NUMBER: PCT/US99/30095  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: PCT/US99/30999  
PRIOR FILING DATE: 1999-12-20  
PRIOR APPLICATION NUMBER: PCT/US00/00376  
PRIOR FILING DATE: 2000-01-06  
PRIOR APPLICATION NUMBER: PCT/US00/03565  
PRIOR FILING DATE: 2000-02-11  
PRIOR APPLICATION NUMBER: PCT/US00/04341  
PRIOR FILING DATE: 2000-02-18  
PRIOR APPLICATION NUMBER: PCT/US00/04342

```
; PRIOR FILING DATE: 2000-02-18
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: PCT/US00/06884
; PRIOR FILING DATE: 2000-03-15
; PRIOR APPLICATION NUMBER: PCT/US00/08439
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: PCT/US00/13705
; PRIOR FILING DATE: 2000-05-17
; PRIOR APPLICATION NUMBER: PCT/US00/14042
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: PCT/US00/14941
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: PCT/US00/15264
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: PCT/US00/22031
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: PCT/US00/23522
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: PCT/US00/30873
; PRIOR FILING DATE: 2000-11-10
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: PCT/US01/06666
; PRIOR FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: PCT/US01/17092
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: PCT/US01/17800
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US01/19692
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: PCT/US01/21066
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: PCT/US01/21735
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: PCT/US01/27099
; PRIOR FILING DATE: 2001-08-29
; NUMBER OF SEQ ID NOS: 91
; SEQ ID NO 20
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-001-054-20

Query Match 1.5%; Score 6; DB 9; Length 92;
Best Local Similarity 100.0%; Pred.No.1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 LSLPLL 15
Db 45 LSLPLL 50

RESULT 42
US-10-052-586-446
; Sequence 446. Application US/10052586
; Patent No. US20020127584A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3430R1C1
; CURRENT APPLICATION NUMBER: US/10/052,586
; CURRENT FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063564
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063734
; PRIOR FILING DATE: 1997-10-29
; PRIOR APPLICATION NUMBER: 60/063870
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/064103
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066120
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: 60/066466
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/066772
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/069335
; PRIOR FILING DATE: 1997-12-11
; PRIOR APPLICATION NUMBER: 60/069425
; PRIOR FILING DATE: 1997-12-12
; PRIOR APPLICATION NUMBER: 60/069870
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/068017
; PRIOR FILING DATE: 1997-12-18
; PRIOR APPLICATION NUMBER: 60/077450
; PRIOR FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/077632
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/077649
; PRIOR FILING DATE: 1998-03-11
; PRIOR APPLICATION NUMBER: 60/078886
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/078939
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079664
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/079786
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/080107
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080194
; PRIOR FILING DATE: 1998-03-31
; PRIOR APPLICATION NUMBER: 60/080327
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080333
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/081049
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081070
; PRIOR FILING DATE: 1998-04-08
; PRIOR APPLICATION NUMBER: 60/081195
; PRIOR FILING DATE: 1998-04-09
```

;; PRIOR APPLICATION NUMBER: 60/081838  
;; PRIOR FILING DATE: 1998-04-15  
;; PRIOR APPLICATION NUMBER: 60/082568  
;; PRIOR FILING DATE: 1998-04-21  
;; PRIOR APPLICATION NUMBER: 60/082569  
;; PRIOR FILING DATE: 1998-04-21  
;; PRIOR APPLICATION NUMBER: 60/082704  
;; PRIOR FILING DATE: 1998-04-22  
;; PRIOR APPLICATION NUMBER: 60/082797  
;; PRIOR FILING DATE: 1998-04-22  
;; PRIOR APPLICATION NUMBER: 60/083322  
;; PRIOR FILING DATE: 1998-04-28  
;; PRIOR APPLICATION NUMBER: 60/083495  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083496  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083499  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/083559  
;; PRIOR FILING DATE: 1998-04-29  
;; PRIOR APPLICATION NUMBER: 60/084366  
;; PRIOR FILING DATE: 1998-05-05  
;; PRIOR APPLICATION NUMBER: 60/084414  
;; PRIOR FILING DATE: 1998-05-06  
;; PRIOR APPLICATION NUMBER: 60/084639  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084640  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/084643  
;; PRIOR FILING DATE: 1998-05-07  
;; PRIOR APPLICATION NUMBER: 60/085573  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085579  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085580  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085582  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/085700  
;; PRIOR FILING DATE: 1998-05-15  
;; PRIOR APPLICATION NUMBER: 60/086023  
;; PRIOR FILING DATE: 1998-05-18  
;; PRIOR APPLICATION NUMBER: 60/086392  
;; PRIOR FILING DATE: 1998-05-22  
;; PRIOR APPLICATION NUMBER: 60/086486  
;; PRIOR FILING DATE: 1998-05-22  
;; PRIOR APPLICATION NUMBER: 60/087098  
;; PRIOR FILING DATE: 1998-05-28  
;; PRIOR APPLICATION NUMBER: 60/087208  
;; PRIOR FILING DATE: 1998-05-28  
;; PRIOR APPLICATION NUMBER: 60/087609  
;; PRIOR FILING DATE: 1998-06-02  
;; PRIOR APPLICATION NUMBER: 60/087759  
;; PRIOR FILING DATE: 1998-06-02  
;; PRIOR APPLICATION NUMBER: 60/087827  
;; PRIOR FILING DATE: 1998-06-03  
;; PRIOR APPLICATION NUMBER: 60/088025  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088028  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088029  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088033  
;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088167  
;; PRIOR FILING DATE: 1998-06-05  
;; PRIOR APPLICATION NUMBER: 60/088202  
;; PRIOR FILING DATE: 1998-06-05  
;; PRIOR APPLICATION NUMBER: 60/088212  
;; PRIOR FILING DATE: 1998-06-05  
;; PRIOR APPLICATION NUMBER: 60/088217  
;; PRIOR FILING DATE: 1998-06-05  
;; PRIOR APPLICATION NUMBER: 60/088326

;; PRIOR FILING DATE: 1998-06-04  
;; PRIOR APPLICATION NUMBER: 60/088655  
;; PRIOR FILING DATE: 1998-06-09  
;; PRIOR APPLICATION NUMBER: 60/088722  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088738  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088740  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088811  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088824  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088825  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088826  
;; PRIOR FILING DATE: 1998-06-10  
;; PRIOR APPLICATION NUMBER: 60/088861  
;; PRIOR FILING DATE: 1998-06-11  
;; PRIOR APPLICATION NUMBER: 60/088863  
;; PRIOR FILING DATE: 1998-06-11  
;; PRIOR APPLICATION NUMBER: 60/088876  
;; PRIOR FILING DATE: 1998-06-11  
;; PRIOR APPLICATION NUMBER: 60/089512  
;; PRIOR FILING DATE: 1998-06-16  
;; PRIOR APPLICATION NUMBER: 60/089514  
;; PRIOR FILING DATE: 1998-06-16  
;; PRIOR APPLICATION NUMBER: 60/089538  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089598  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089653  
;; PRIOR FILING DATE: 1998-06-17  
;; PRIOR APPLICATION NUMBER: 60/089908

Query Match 1.5%; Score 6; DB 12; Length 92;  
Best Local Similarity 100.0%; Pred.No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 10 LSLPL 15  
Db 45 LSLPL 50

RESULT 43  
US-09-981-876-203  
; Sequence 203; Application US/09981876  
; Patent No. US2002016469A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 70 Human Secreted Proteins  
; FILE REFERENCE: P2001P1  
; CURRENT APPLICATION NUMBER: US/09/981,876  
; CURRENT FILING DATE: 2001-10-19  
; PRIOR APPLICATION NUMBER: 09/148,545  
; PRIOR FILING DATE: 1998-09-04  
; PRIOR APPLICATION NUMBER: 60/040,162  
; PRIOR FILING DATE: 1997-03-07  
; PRIOR APPLICATION NUMBER: 60/040,333  
; PRIOR FILING DATE: 1997-03-07  
; PRIOR APPLICATION NUMBER: 60/038,621  
; PRIOR FILING DATE: 1997-03-07  
; PRIOR APPLICATION NUMBER: 60/040,161  
; PRIOR FILING DATE: 1997-03-07  
; PRIOR APPLICATION NUMBER: 60/040,626  
; PRIOR FILING DATE: 1997-03-07  
; PRIOR APPLICATION NUMBER: 60/040,334  
; PRIOR FILING DATE: 1997-03-07  
; PRIOR APPLICATION NUMBER: 60/040,336

1	PRIOR APPLICATION NUMBER: 60/056,888
2	PRIOR FILING DATE: 1997-08-22
3	PRIOR APPLICATION NUMBER: 60/056,877
4	PRIOR FILING DATE: 1997-08-22
5	PRIOR APPLICATION NUMBER: 60/056,889
6	PRIOR FILING DATE: 1997-08-22
7	PRIOR APPLICATION NUMBER: 60/056,893
8	PRIOR FILING DATE: 1997-08-22
9	PRIOR APPLICATION NUMBER: 60/056,630
10	PRIOR FILING DATE: 1997-08-22
11	PRIOR APPLICATION NUMBER: 60/056,878
12	PRIOR FILING DATE: 1997-08-22
13	PRIOR APPLICATION NUMBER: 60/056,662
14	PRIOR FILING DATE: 1997-08-22
15	PRIOR APPLICATION NUMBER: 60/056,872
16	PRIOR FILING DATE: 1997-08-22
17	PRIOR APPLICATION NUMBER: 60/056,882
18	PRIOR FILING DATE: 1997-08-22
19	PRIOR APPLICATION NUMBER: 60/056,637
20	PRIOR FILING DATE: 1997-08-22
21	PRIOR APPLICATION NUMBER: 60/056,903
22	PRIOR FILING DATE: 1997-08-22
23	PRIOR APPLICATION NUMBER: 60/056,888
24	PRIOR FILING DATE: 1997-08-22
25	PRIOR APPLICATION NUMBER: 60/056,879
26	PRIOR FILING DATE: 1997-08-22
27	PRIOR APPLICATION NUMBER: 60/056,880
28	PRIOR FILING DATE: 1997-08-22
29	PRIOR APPLICATION NUMBER: 60/056,894
30	PRIOR FILING DATE: 1997-08-22
31	PRIOR APPLICATION NUMBER: 60/056,911
32	PRIOR FILING DATE: 1997-08-22
33	PRIOR APPLICATION NUMBER: 60/056,636
34	PRIOR FILING DATE: 1997-08-22
35	PRIOR APPLICATION NUMBER: 60/056,864
36	PRIOR FILING DATE: 1997-08-22
37	PRIOR APPLICATION NUMBER: 60/056,631
38	PRIOR FILING DATE: 1997-08-22
39	PRIOR APPLICATION NUMBER: 60/056,845
40	PRIOR FILING DATE: 1997-08-22
41	PRIOR APPLICATION NUMBER: 60/056,892
42	PRIOR FILING DATE: 1997-08-22
43	PRIOR APPLICATION NUMBER: 60/047,595
44	PRIOR FILING DATE: 1997-05-23
45	PRIOR APPLICATION NUMBER: 60/057,761
46	PRIOR FILING DATE: 1997-05-23
47	PRIOR APPLICATION NUMBER: 60/047,585
48	PRIOR FILING DATE: 1997-05-23
49	PRIOR APPLICATION NUMBER: 60/047,594
50	PRIOR FILING DATE: 1997-05-23
51	PRIOR APPLICATION NUMBER: 60/047,589
52	PRIOR FILING DATE: 1997-05-23
53	PRIOR APPLICATION NUMBER: 60/047,588
54	PRIOR FILING DATE: 1997-05-23
55	PRIOR APPLICATION NUMBER: 60/047,585
56	PRIOR FILING DATE: 1997-05-23
57	PRIOR APPLICATION NUMBER: 60/047,594
58	PRIOR FILING DATE: 1997-05-23
59	PRIOR APPLICATION NUMBER: 60/043,578
60	PRIOR FILING DATE: 1997-04-11
61	PRIOR APPLICATION NUMBER: 60/043,576
62	PRIOR FILING DATE: 1997-04-11
63	PRIOR APPLICATION NUMBER: 60/047,501
64	PRIOR FILING DATE: 1997-04-11



PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/043,670  
PRIOR FILING DATE: 1997-04-11  
PRIOR APPLICATION NUMBER: 60/056,632  
PRIOR FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: 60/056,664  
PRIOR FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: 60/056,876  
PRIOR FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: 60/056,881  
PRIOR FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: 60/056,909  
PRIOR FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: 60/056,875  
PRIOR FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: 60/056,862  
PRIOR FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: 60/056,887  
PRIOR FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: 60/056,908  
PRIOR FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: 60/048,964  
PRIOR FILING DATE: 1997-06-06  
PRIOR APPLICATION NUMBER: 60/057,650  
PRIOR FILING DATE: 1997-09-05  
PRIOR APPLICATION NUMBER: 60/056,884  
PRIOR FILING DATE: 1997-08-22  
NUMBER OF SEQ ID NOS: 260  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 203  
LENGTH: 93

Query Match 1.5%; Score 6; DB 9; Length 93;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 10 LSLPL 15  
Db 45 LSLPL 50

RESULT 44  
US-09-981-876-256  
Sequence 256, Application US/09981876  
Patent No. US20020164669A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OR INVENTION: 70 Human Secreted Proteins  
FILE REFERENCE: P2001P1  
CURRENT APPLICATION NUMBER: US/09/981,876  
PRIOR FILING DATE: 2001-10-19  
PRIOR APPLICATION NUMBER: 09/148,545  
PRIOR FILING DATE: 1998-09-04  
PRIOR APPLICATION NUMBER: 60/040,162  
PRIOR FILING DATE: 1997-03-07  
PRIOR APPLICATION NUMBER: 60/040,333  
PRIOR FILING DATE: 1997-03-07  
PRIOR APPLICATION NUMBER: 60/038,621  
PRIOR FILING DATE: 1997-03-07  
PRIOR APPLICATION NUMBER: 60/040,161  
PRIOR FILING DATE: 1997-03-07  
PRIOR APPLICATION NUMBER: 60/040,626  
PRIOR FILING DATE: 1997-03-07  
PRIOR APPLICATION NUMBER: 60/040,334  
PRIOR FILING DATE: 1997-03-07  
PRIOR APPLICATION NUMBER: 60/040,336  
PRIOR FILING DATE: 1997-03-07  
PRIOR APPLICATION NUMBER: 60/040,163  
PRIOR FILING DATE: 1997-03-07  
PRIOR APPLICATION NUMBER: 60/047,615  
PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/047,600  
PRIOR FILING DATE: 1997-05-23

PRIOR APPLICATION NUMBER: 60/047,597  
PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/047,502  
PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/047,633  
PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/047,583  
PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/047,617  
PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/047,618  
PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/047,503  
PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/047,592  
PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/047,581  
PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/047,584  
PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/047,500  
PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/047,587  
PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/047,492  
PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/047,598  
PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/047,613  
PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/047,582  
PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/047,596  
PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/047,612  
PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/047,632  
PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/047,601  
PRIOR FILING DATE: 1997-05-23  
PRIOR APPLICATION NUMBER: 60/043,580  
PRIOR FILING DATE: 1997-04-11  
PRIOR APPLICATION NUMBER: 60/043,568  
PRIOR FILING DATE: 1997-04-11  
PRIOR APPLICATION NUMBER: 60/043,314  
PRIOR FILING DATE: 1997-04-11  
PRIOR APPLICATION NUMBER: 60/043,569  
PRIOR FILING DATE: 1997-04-11  
PRIOR APPLICATION NUMBER: 60/043,311  
PRIOR FILING DATE: 1997-04-11  
PRIOR APPLICATION NUMBER: 60/043,671  
PRIOR FILING DATE: 1997-04-11  
PRIOR APPLICATION NUMBER: 60/043,674  
PRIOR FILING DATE: 1997-04-11  
PRIOR APPLICATION NUMBER: 60/043,669  
PRIOR FILING DATE: 1997-04-11  
PRIOR APPLICATION NUMBER: 60/043,312  
PRIOR FILING DATE: 1997-04-11  
PRIOR APPLICATION NUMBER: 60/043,313  
PRIOR FILING DATE: 1997-04-11  
PRIOR APPLICATION NUMBER: 60/043,672  
PRIOR FILING DATE: 1997-04-11  
PRIOR APPLICATION NUMBER: 60/043,315  
PRIOR FILING DATE: 1997-04-11  
PRIOR APPLICATION NUMBER: 60/048,974  
PRIOR FILING DATE: 1997-06-06  
PRIOR APPLICATION NUMBER: 60/056,886  
PRIOR FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: 60/056,877  
PRIOR FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: 60/056,889  
PRIOR FILING DATE: 1997-08-22  
PRIOR APPLICATION NUMBER: 60/056,893

; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,630  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,878  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,662  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,872  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,882  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,637  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,903  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,888  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,879  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,880  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,894  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,911  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,636  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,874  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,910  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,864  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,631  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,845  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,892  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/047,595  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/057,761  
; PRIOR FILING DATE: 05-Sep-1997  
; PRIOR APPLICATION NUMBER: 60/047,599  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/047,588  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/047,585  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/047,586  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/047,590  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/047,594  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/047,589  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/047,593  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/047,614  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/043,578  
; PRIOR FILING DATE: 1997-04-11  
; PRIOR APPLICATION NUMBER: 60/043,576  
; PRIOR FILING DATE: 1997-04-11  
; PRIOR APPLICATION NUMBER: 60/047,501  
; PRIOR FILING DATE: 1997-05-23  
; PRIOR APPLICATION NUMBER: 60/043,670  
; PRIOR FILING DATE: 1997-04-11  
; PRIOR APPLICATION NUMBER: 60/056,632  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,664  
; PRIOR FILING DATE: 1997-08-22

; PRIOR APPLICATION NUMBER: 60/056,876  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,881  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,909  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,875  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,862  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,887  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/056,908  
; PRIOR FILING DATE: 1997-08-22  
; PRIOR APPLICATION NUMBER: 60/048,964  
; PRIOR FILING DATE: 1997-06-06  
; PRIOR APPLICATION NUMBER: 60/057,650  
; PRIOR FILING DATE: 1997-09-05  
; PRIOR APPLICATION NUMBER: 60/056,884  
; PRIOR FILING DATE: 1997-08-22  
; NUMBER OF SEQ ID NOS: 280  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 256  
; LENGTH: 93

Query Match 1.5%; Score 6; DB 9; Length 93;  
Best Local Similarity 100.0%; Pred.No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 10 LSLPLL 15  
Db 45 LSLPLL 50

RESULT 45  
US-09-864-761-48467  
; Sequence 48467, Application US/09864761  
; Patent NO. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aesomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 48467
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000012.2
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
; OTHER INFORMATION: EST HUMAN HIT: AUI19258.1, EVALUE 2.00e-29
; OTHER INFORMATION: SWISSPROT HIT: Q63470, EVALUE 2.00e-30
US-09-864-761-48467

```

```

Query Match          1.5%; Score 6; DB 10; Length 95;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 93 SSGTTS 98
Db 84 SSGTTS 89

```

Search completed: December 25, 2002, 20:36:30  
 Job time : 27 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 25, 2002, 20:28:58 ; Search time 19 Seconds  
(without alignments)  
2023.883 Million cell updates/sec

Title: US-09-889-746-2

Perfect score: 400  
Sequence: 1 MRSHYFQWLSPLLSVAVT.....SKAEVLGRVPLNKPKRALF 400

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	2.5	725	2 S01042	aerobactin recepto
2	10	2.5	726	2 AB0122	receptor-like prot
3	10	2.5	732	2 T44483	probable sideropho
4	9	2.2	813	2 P83476	probable integrin
5	8	2.0	334	2 T29061	short neurotoxin I
6	7	1.8	62	2 B25866	short neurotoxin d
7	7	1.8	62	2 D25866	short neurotoxin c
8	7	1.8	62	2 AG0748	hypothetical prote
9	7	1.8	63	2 G86631	hypothetical prote
10	7	1.8	87	2 H86382	hypothetical prote
11	7	1.8	131	2 T18204	mxh protein - Bac
12	7	1.8	154	2 T18204	vasotocin - Pacifi
13	7	1.8	161	2 B84348	hypothetical prote
14	7	1.8	173	2 B87268	grip protein (lipo
15	7	1.8	198	2 H83059	hypothetical prote
16	7	1.8	200	2 AE2461	hypothetical prote
17	7	1.8	223	2 T36587	hypothetical prote
18	7	1.8	225	2 H82789	conserved hypothet
19	7	1.8	233	2 DSHUEC	superoxide dismuta
20	7	1.8	240	1 B83799	hypothetical prote
21	7	1.8	249	2 B90421	hypothetical prote
22	7	1.8	254	2 T33405	hypothetical prote
23	7	1.8	260	2 A82304	3-methyl-2-oxobuta
24	7	1.8	271	2 B98179	probable NAD(P)H n
25	7	1.8	275	1 C69030	MU1225 protein hom
26	7	1.8	281	2 G86269	protein F21F23.9 f
27	7	1.8	290	2 E85709	unknown protein en
28	7	1.8	293	2 A89618	hypothetical prote
29	7	1.8	293	2 A89618	hypothetical prote

30	7	1.8	325	2 A13635	thiamine biosynthe
31	7	1.8	334	2 H95342	Nox protein requi
32	7	1.8	335	2 B97688	C4-dicarboxylate c
33	7	1.8	335	2 AG2913	C4-dicarboxylate b
34	7	1.8	337	2 C82190	formate dehydrogen
35	7	1.8	340	2 B75436	conserved hypothet
36	7	1.8	349	2 AH3338	cobw protein (lipo
37	7	1.8	354	2 AD0374	phosphoribosylamin
38	7	1.8	361	2 B70189	rod shape-determin
39	7	1.8	361	2 E96904	mind family ATPase
40	7	1.8	362	2 AP1970	3-isopropylmalate
41	7	1.8	362	2 A41601	Na+/taurocholate c
42	7	1.8	369	2 C90646	probable fibribial
43	7	1.8	369	2 C85497	probable fibribial
44	7	1.8	380	1 GNVSMB	genome polyprotein
45	7	1.8	385	2 B90682	probable enzyme (i
46	7	1.8	385	2 P85532	probable enzyme ya
47	7	1.8	385	2 AP0548	penicillin-binding
48	7	1.8	385	2 H64765	yaH protein - Bac
49	7	1.8	386	2 D71851	tetracycline resis
50	7	1.8	386	2 B64665	tetracycline resis
51	7	1.8	390	2 AP0426	probable GTP-bindi
52	7	1.8	399	1 HMKRHZ	sigma 1 protein -
53	7	1.8	425	2 C87619	hypothetical prote
54	7	1.8	430	2 D72324	O-acetylhomoserine
55	7	1.8	433	2 C44465	sodium ion pump ox
56	7	1.8	433	2 AD0909	oxaloacetate decar
57	7	1.8	433	2 A10508	mannose-6-phosphat
58	7	1.8	453	2 H70352	late expression fa
59	7	1.8	457	2 T10360	sigma 1 protein -
60	7	1.8	462	2 C34829	acid phosphatase (
61	7	1.8	466	1 T04599	flagellar hook-aaa
62	7	1.8	466	2 AF0224	ubiquinol-cytochro
63	7	1.8	474	2 JX0301	hypothetical prote
64	7	1.8	508	2 B91250	hypothetical prote
65	7	1.8	519	2 T48498	hypothetical prote
66	7	1.8	527	2 T22559	hypothetical prote
67	7	1.8	538	2 P83622	probable permease
68	7	1.8	546	2 F71951	periplasmic dipept
69	7	1.8	563	2 C95874	conserved hypothet
70	7	1.8	576	2 A71497	probable DNA mima
71	7	1.8	580	2 A83874	carbon starvation-
72	7	1.8	587	2 G64624	outer membrane pro
73	7	1.8	587	2 C71889	probable outer mem
74	7	1.8	590	2 I40608	cspa protein - Clo
75	7	1.8	617	2 S74002	translation initia
76	7	1.8	648	1 W0EC2N	phosphotransferase
77	7	1.8	648	2 B90717	hypothetical prote
78	7	1.8	658	2 B85567	hypothetical prote
79	7	1.8	665	2 T15801	hypothetical prote
80	7	1.8	668	2 B89783	hypothetical prote
81	7	1.8	674	2 T48261	hypothetical prote
82	7	1.8	717	2 B97437	glycyl-tRNA synthe
83	7	1.8	717	2 AF2655	hypothetical prote
84	7	1.8	726	2 T31287	hypothetical prote
85	7	1.8	745	2 T07184	subtilisin-like pr
86	7	1.8	745	2 J06119	subtilisin-like pr
87	7	1.8	746	2 T46821	subtilisin-like pr
88	7	1.8	746	2 A95420	siderophore recept
89	7	1.8	747	2 T06580	RhaR Rhizobactin r
90	7	1.8	754	2 T06577	subtilisin-like pr
91	7	1.8	754	2 T06579	subtilisin-like pr
92	7	1.8	793	1 SUPFCA	subtilisin-like pr
93	7	1.8	803	2 B82045	aspartokinase II/h
94	7	1.8	804	2 T18014	hypothetical prote
95	7	1.8	808	2 C72858	AcOrf-66 protein -
96	7	1.8	820	2 T04227	hypothetical prote
97	7	1.8	867	2 AC2140	ferrichrome-iron r
98	7	1.8	1029	2 H66179	hypothetical prote
99	7	1.8	1049	2 T42045	beta transducin-li
100	7	1.8	1051	2 T05426	hypothetical prote
101	7	1.8	1089	2 T30843	serine-repeat anti
102	7	1.8	1125	2 H87644	TonB-dependent rec

103	7	1.8	1157	2	A55152	PAS1 protein - yea	176	6	1.5	131	2	AP1944	arsenate reductase
104	7	1.8	1206	2	T34021	protein kinase SK2	177	6	1.5	131	2	S36301	T-cell receptor de
105	7	1.8	1215	2	E70814	hypothetical prote	178	6	1.5	131	2	T08788	hypothetical prote
106	7	1.8	1381	2	E70806	hypothetical glyci	179	6	1.5	135	2	F70639	hypothetical prote
107	7	1.8	1969	2	T38495	hypothetical prote	180	6	1.5	135	2	T46252	hypothetical prote
108	6	1.5	33	2	A46116	penton base protei	181	6	1.5	136	2	S36304	T-cell receptor de
109	6	1.5	45	2	A39598	TAP-29 anti-HIV pr	182	6	1.5	136	2	AG2009	transcription regu
110	6	1.5	45	2	A40432	calpain IEC 3.4.22	183	6	1.5	137	2	S36303	T-cell receptor de
111	6	1.5	46	2	T31632	hypothetical prote	184	6	1.5	137	2	AB2544	transcription regu
112	6	1.5	46	2	G89761	hypothetical prote	185	6	1.5	137	2	AC0719	probable pertuasis
113	6	1.5	53	2	E95311	hypothetical prote	186	6	1.5	138	2	S06950	Pox meso protein -
114	6	1.5	53	2	AG1549	hypothetical prote	187	6	1.5	138	2	S78252	ribosomal protein
115	6	1.5	53	2	AI1191	hypothetical prote	188	6	1.5	139	2	S36302	T-cell receptor de
116	6	1.5	55	2	A83500	periplasmic nitrat	189	6	1.5	139	2	FC1213	Ig heavy chain pre
117	6	1.5	55	2	G43256	hypothetical prote	190	6	1.5	139	2	G29380	Ig heavy chain pre
118	6	1.5	56	1	ERBP1B	gene 1B protein -	191	6	1.5	140	2	A29902	paired box homolog
119	6	1.5	57	2	D44530	T-cell receptor al	192	6	1.5	140	2	T10589	hypothetical prote
120	6	1.5	61	2	B96577	probable RNA polym	193	6	1.5	140	2	AD2774	conserved hypotet
121	6	1.5	62	2	D86516	hypothetical prote	194	6	1.5	140	2	B97554	hypothetical prote
122	6	1.5	62	2	E72105	hypothetical prote	195	6	1.5	141	2	S36318	T-cell receptor de
123	6	1.5	66	2	T44533	hypothetical prote	196	6	1.5	141	2	A44464	oxaloacetate decar
124	6	1.5	66	2	E40361	virC-region hypoth	197	6	1.5	142	2	S36316	T-cell receptor de
125	6	1.5	70	2	T44130	hypothetical prote	198	6	1.5	143	2	S36321	T-cell receptor de
126	6	1.5	71	2	G87502	hypothetical prote	199	6	1.5	143	2	D97319	transcription regu
127	6	1.5	76	2	T13117	protein gp31 - pha	200	6	1.5	144	2	B70360	arsenate reductase
128	6	1.5	77	2	AD2662	conserved hypotet	201	6	1.5	144	2	F90021	50S ribosomal prot
129	6	1.5	83	2	S10706	urotensin II precu	202	6	1.5	144	2	T46379	hypothetical prote
130	6	1.5	86	1	Q1BP67	gene 1.6 protein -	203	6	1.5	145	1	R3YU16	ribosomal protein
131	6	1.5	87	2	S36319	T-cell receptor de	204	6	1.5	145	2	A36569	plastoocyanin precu
132	6	1.5	87	2	AH3300	hypothetical cytos	205	6	1.5	145	2	S36299	T-cell receptor de
133	6	1.5	91	2	E90766	hypothetical prote	206	6	1.5	145	2	S41193	ribosomal protein
134	6	1.5	91	2	F85716	unknown protein en	207	6	1.5	145	2	S34574	gene 70 protein -
135	6	1.5	92	2	E87002	conserved hypotet	208	6	1.5	145	2	AE1393	wall teichoic acid
136	6	1.5	93	2	I50673	gene Pax-1 protein	209	6	1.5	145	2	AH1768	hypothetical prote
137	6	1.5	94	2	A84966	integration host f	210	6	1.5	145	2	E97694	hypothetical prote
138	6	1.5	94	2	T11595	hypothetical prote	211	6	1.5	145	2	AD2920	cobalamin biosynth
139	6	1.5	94	2	C89948	hypothetical prote	212	6	1.5	146	2	E84489	40S ribosomal prot
140	6	1.5	95	2	AD1168	hypothetical prote	213	6	1.5	147	2	F72091	type III secretion
141	6	1.5	96	1	JU0349	11.5K protein - Ch	214	6	1.5	147	2	E86531	secretion chaparon
142	6	1.5	97	1	S16028	Ig heavy chain V r	215	6	1.5	148	2	T20053	hypothetical prote
143	6	1.5	97	2	B96924	hypothetical prote	216	6	1.5	149	2	S36317	T-cell receptor de
144	6	1.5	98	2	T28629	Y4CB protein - Rhi	217	6	1.5	150	1	G69873	yhcv-related prote
145	6	1.5	100	2	C97444	hypothetical prote	218	6	1.5	150	2	D72670	hypothetical prote
146	6	1.5	105	2	S66917	hypothetical prote	219	6	1.5	151	1	R5N728	ribosomal protein
147	6	1.5	105	2	T48733	hypothetical prote	220	6	1.5	152	2	S57421	cysteine proteinas
148	6	1.5	109	2	S77796	phosphotransferase	221	6	1.5	152	2	C69546	hypothetical prote
149	6	1.5	109	2	G71249	hypothetical prote	222	6	1.5	153	2	A64998	Protein elaa - Esc
150	6	1.5	109	2	AH3240	conserved hypotet	223	6	1.5	153	2	C91023	hypothetical prote
151	6	1.5	111	2	S51032	subtilisin-chymotr	224	6	1.5	153	2	D85867	hypothetical prote
152	6	1.5	112	2	B95024	conserved domain p	225	6	1.5	154	2	G64023	hypothetical prote
153	6	1.5	114	2	T37050	probable integral	226	6	1.5	154	2	D83203	hypothetical prote
154	6	1.5	117	2	H83815	hypothetical prote	227	6	1.5	155	2	C97383	hypothetical prote
155	6	1.5	117	2	AH2182	hypothetical prote	228	6	1.5	156	2	AB2601	conserved hypotet
156	6	1.5	118	1	Q0BE8H	hypothetical 13K p	229	6	1.5	156	2	D72463	hypothetical prote
157	6	1.5	118	2	H71564	hypothetical prote	230	6	1.5	156	2	AP1805	hypothetical prote
158	6	1.5	120	2	T16502	hypothetical prote	231	6	1.5	156	2	AH1431	hypothetical prote
159	6	1.5	120	2	B41061	paired box homolog	232	6	1.5	157	2	G75266	hypothetical prote
160	6	1.5	120	2	H82831	preprotein tranalo	233	6	1.5	157	2	T72535	hypothetical prote
161	6	1.5	120	2	F71126	hypothetical prote	234	6	1.5	157	2	T49394	hypothetical prote
162	6	1.5	120	4	Q0HUPL	hypothetical prote	235	6	1.5	157	2	B47150	high mobility grou
163	6	1.5	121	2	B36858	GIL protein - vari	236	6	1.5	159	2	AD1199	methylyated-DNA-pro
164	6	1.5	122	2	G82217	probable 6-pyruvoy	237	6	1.5	159	2	A72462	hypothetical prote
165	6	1.5	123	2	C87698	conserved hypotet	238	6	1.5	159	2	T05279	transcription fact
166	6	1.5	124	2	S40099	hypothetical prote	239	6	1.5	160	2	A48975	thy-1 glycoprotein
167	6	1.5	124	2	D71207	hypothetical prote	240	6	1.5	160	2	S58759	ezrin - rat (fragm
168	6	1.5	124	2	T36629	probable transcrip	241	6	1.5	161	2	F75329	conserved hypotet
169	6	1.5	125	2	S40315	Ig kappa chain - h	242	6	1.5	161	2	F89888	hypothetical prote
170	6	1.5	126	2	B99993	hypothetical prote	243	6	1.5	161	2	S18991	high mobility grou
171	6	1.5	127	2	T26408	hypothetical prote	244	6	1.5	162	2	G64351	hypothetical prote
172	6	1.5	127	2	A72670	hypothetical prote	245	6	1.5	162	2	T21832	hypothetical prote
173	6	1.5	128	1	AZPSDF	azurin - Pseudomon	246	6	1.5	162	2	D97821	hypothetical prote
174	6	1.5	130	2	I39704	hypothetical prote	247	6	1.5	164	2	T36471	hypothetical prote
175	6	1.5	131	1	RDEB15	fumurate reductase	248	6	1.5	165	2	B56236	probable RNA helic

249	6	1.5	167	2	A95346	NapB periplasmic n	322	6	1.5	206	2	JN0623	GRP-binding protei
250	6	1.5	168	2	T03640	high mobility grou	323	6	1.5	206	2	B65093	hypothetical prote
251	6	1.5	169	2	H90970	probable terminase	324	6	1.5	206	2	B91121	hypothetical prote
252	6	1.5	169	2	G85743	probable terminase	325	6	1.5	206	2	A85966	hypothetical prote
253	6	1.5	169	2	H72470	hypothetical prote	326	6	1.5	207	2	AP0953	glucose inhibited
254	6	1.5	169	2	H84191	hypothetical prote	327	6	1.5	207	2	AF0080	probable exported
255	6	1.5	170	1	B47673	flavodoxin 1818 -	328	6	1.5	207	2	D87267	hypothetical prote
256	6	1.5	170	2	S56958	probable membrane	329	6	1.5	207	2	S20683	phosphomethylmal
257	6	1.5	171	2	A12594	hypothetical prote	330	6	1.5	208	2	S36155	paired box protein
258	6	1.5	172	2	AF3496	3-hydroxydecanoyl-	331	6	1.5	209	2	T36702	probable transcrip
259	6	1.5	172	2	AD0451	conserved hypothet	332	6	1.5	209	2	JE0154	mitochondrial inne
260	6	1.5	174	2	C84066	hypothetical prote	333	6	1.5	210	2	AE4108	glycerol-3-phospha
261	6	1.5	175	2	D82737	hypothetical prote	334	6	1.5	212	2	S06961	gene HUP48 protein
262	6	1.5	175	2	F86903	cell shape determi	335	6	1.5	212	2	C70898	hypothetical prote
263	6	1.5	176	2	E87523	hypoxanthine-guanl	336	6	1.5	212	2	E75339	hypothetical prote
264	6	1.5	176	2	T03420	crsf protein - Agr	337	6	1.5	214	2	S29793	protein ORF 214 (a
265	6	1.5	176	2	AP3243	conjugal transfer	338	6	1.5	214	2	AC2041	hypothetical prote
266	6	1.5	176	2	H95272	hypothetical prote	339	6	1.5	215	2	S61841	adenylate kinase (
267	6	1.5	178	2	AE4115	parB protein homol	340	6	1.5	215	2	S61843	adenylate kinase (
268	6	1.5	178	2	T51159	HMG protein (impor	341	6	1.5	215	2	F81154	transaldolase, TAL
269	6	1.5	179	2	B64335	hypothetical prote	342	6	1.5	215	2	H97065	hypothetical prote
270	6	1.5	180	2	T33142	hypothetical prote	343	6	1.5	215	2	S64438	hypothetical prote
271	6	1.5	180	2	C83325	hypothetical prote	344	6	1.5	216	2	C75102	hypothetical prote
272	6	1.5	180	2	T06135	hypothetical prote	345	6	1.5	216	2	T18176	hypothetical prote
273	6	1.5	181	2	H66745	probable blue copp	346	6	1.5	217	1	G70459	conserved hypothet
274	6	1.5	182	2	S28310	hypothetical prote	347	6	1.5	217	2	AC2428	hypothetical prote
275	6	1.5	182	2	S69469	hypothetical prote	348	6	1.5	219	2	H89811	hypothetical prote
276	6	1.5	183	2	B69353	hypothetical prote	349	6	1.5	220	2	B84241	carbonic anhydrase
277	6	1.5	183	2	D64795	Apo-citrate lyase	350	6	1.5	220	2	T39548	hypothetical prote
278	6	1.5	183	2	A85561	hypothetical prote	351	6	1.5	221	2	C81842	hypothetical prote
279	6	1.5	183	2	B90710	Apo-citrate lyase	352	6	1.5	221	2	H82999	hypothetical prote
280	6	1.5	185	2	C97377	3-hydroxydecanoyl-	353	6	1.5	222	2	C70980	probable hydrolase
281	6	1.5	185	2	D90863	hypothetical prote	354	6	1.5	222	2	S36139	hypothetical prote
282	6	1.5	185	2	E85755	hypothetical prote	355	6	1.5	222	2	E90321	mitochondrial inne
283	6	1.5	185	2	F64878	aldehyde dehydroge	356	6	1.5	222	2	T37019	hypothetical prote
284	6	1.5	186	2	D69007	ribosomal protein	357	6	1.5	224	2	E71319	probable integral
285	6	1.5	186	2	H86357	F12X8.18 protein -	358	6	1.5	224	2	H86117	hypothetical prote
286	6	1.5	187	2	A90893	probable type 1 fl	359	6	1.5	224	2	H91276	hypothetical prote
287	6	1.5	187	2	T18844	hypothetical prote	360	6	1.5	225	2	B82073	ribonuclease III V
288	6	1.5	187	2	H85724	probable major flm	361	6	1.5	225	2	T39239	hypothetical prote
289	6	1.5	188	2	F64336	dTMP kinase (EC 2.	362	6	1.5	225	2	T04066	hypothetical prote
290	6	1.5	189	2	B86193	hypothetical prote	363	6	1.5	227	2	I37021	cytochrome-c oxida
291	6	1.5	190	2	S28605	GRP-binding protei	364	6	1.5	227	2	T33766	cytochrome-c oxida
292	6	1.5	191	2	T29502	hypothetical prote	365	6	1.5	227	2	T11183	hypothetical prote
293	6	1.5	191	2	AE0042	probable TetR-fam1	366	6	1.5	227	2	T35395	probable transfera
294	6	1.5	193	2	S67659	probable membrane	367	6	1.5	227	2	AF2041	hypothetical prote
295	6	1.5	194	2	C72554	hypothetical prote	368	6	1.5	228	2	B38941	cytochrome-c oxida
296	6	1.5	194	2	S75213	hypothetical prote	369	6	1.5	228	2	B45170	cytochrome-c oxida
297	6	1.5	195	2	A97332	transcription regu	370	6	1.5	228	2	T11091	cytochrome-c oxida
298	6	1.5	197	2	AD2190	hypothetical prote	371	6	1.5	228	2	D90620	cytochrome c oxida
299	6	1.5	197	2	AB2162	hypothetical prote	372	6	1.5	228	2	B72655	probable transcrip
300	6	1.5	198	2	S55131	hypothetical prote	373	6	1.5	229	2	H45170	cytochrome-c oxida
301	6	1.5	199	2	S50717	probable membrane	374	6	1.5	229	2	D35116	hypothetical prote
302	6	1.5	199	2	T39861	mlo3 protein - fls	375	6	1.5	229	2	H83892	two-component resp
303	6	1.5	202	2	B87340	transcription regu	376	6	1.5	229	2	H83183	hypothetical prote
304	6	1.5	202	2	H70760	probable lipoprote	377	6	1.5	232	2	A48384	beta-casein - pig
305	6	1.5	203	2	AC3318	sodium-dependent p	378	6	1.5	233	2	T44518	aquaporin homolog
306	6	1.5	203	2	S00747	somatotropin precu	379	6	1.5	233	2	D95860	probable transcrip
307	6	1.5	203	2	S67607	probable membrane	380	6	1.5	235	2	B86821	glucosamine-6-phos
308	6	1.5	204	1	STF1	somatotropin precu	381	6	1.5	235	2	B86821	integral membrane
309	6	1.5	204	2	I51289	somatotropin - Mor	382	6	1.5	236	1	AE3344	L-fucose-phospha
310	6	1.5	204	2	S01746	somatotropin precu	383	6	1.5	236	2	B97436	hypothetical prote
311	6	1.5	204	2	JH0577	somatotropin precu	384	6	1.5	237	2	D71287	hypothetical prote
312	6	1.5	204	2	S30491	somatotropin - Aca	385	6	1.5	237	2	T25877	hypothetical prote
313	6	1.5	204	2	A56904	somatotropin precu	386	6	1.5	239	2	S64573	hypothetical prote
314	6	1.5	204	2	JC4261	somatotropin precu	387	6	1.5	239	2	B84249	hypothetical prote
315	6	1.5	204	2	AB0892	probable membrane	388	6	1.5	240	2	JE0011	DNA-directed RNA p
316	6	1.5	204	2	A54560	TPA-induced protei	389	6	1.5	240	2	H65935	probable GntR-fam1
317	6	1.5	205	2	F87151	conserved hypothet	390	6	1.5	242	2	A10941	hypothetical prote
318	6	1.5	205	2	B31906	hypothetical prote	391	6	1.5	242	2	A71972	hypothetical prote
319	6	1.5	205	2	B97895	hypothetical prote	392	6	1.5	242	2	C64537	conserved hypothet
320	6	1.5	206	1	TVH0AB	transforming prote	393	6	1.5	242	2	B70366	hypothetical prote
321	6	1.5	206	2	D75560	chymidylate kinase	394	6	1.5	242	2	B83738	hypothetical prote

395	6	1.5	242	2	E81337	ABC-transporter AT	468	6	1.5	266	2	AF2654	conserved hypothet
396	6	1.5	243	2	D70355	conserved hypothet	469	6	1.5	267	2	T26852	hypothetical prote
397	6	1.5	244	2	D70355	SAW-dependent meth	470	6	1.5	267	2	T05014	hypothetical prote
398	6	1.5	245	1	KYBOA	chymotrypsin (EC 3	471	6	1.5	267	2	A71871	hypothetical prote
399	6	1.5	245	2	AG3436	short-chain dehydr	472	6	1.5	267	2	C75131	hypothetical prote
400	6	1.5	245	2	C95061	response regulator	473	6	1.5	268	2	T36747	probable secreted
401	6	1.5	245	2	A97129	ribonuclease HI [i	474	6	1.5	268	2	H82757	hypothetical prote
402	6	1.5	245	2	G79229	response regulator	475	6	1.5	268	2	A71066	hypothetical prote
403	6	1.5	246	2	T51631	probable transcrip	476	6	1.5	269	2	D41317	probable lipopolys
404	6	1.5	246	2	D89851	hypothetical prote	477	6	1.5	269	2	AC0973	lipopolysaccharide
405	6	1.5	246	2	G90039	hypothetical prote	478	6	1.5	269	2	A56487	signal recognition
406	6	1.5	247	2	JU0393	hypothetical prote	479	6	1.5	270	2	C72073	metal dependent hy
407	6	1.5	247	2	JCS032	karasurin - Mongol	480	6	1.5	270	2	C86550	metal dependent hy
408	6	1.5	247	2	F93060	conserved hypothet	481	6	1.5	271	2	F87215	probable hydrolase
409	6	1.5	248	2	T46708	formate acetyltran	482	6	1.5	272	2	D83990	lactose transport
410	6	1.5	248	2	AC1613	pyruvate-formate l	483	6	1.5	272	2	AE0308	conserved hypothet
411	6	1.5	248	2	AG1250	pyruvate-formate l	484	6	1.5	272	2	T19537	hypothetical prote
412	6	1.5	248	2	S10623	tropomyosin 4, fib	485	6	1.5	273	2	S24044	lactin precursor -
413	6	1.5	248	2	JCS479	hypothetical prote	486	6	1.5	273	2	B83318	hypothetical prote
414	6	1.5	249	2	F93524	conserved hypothet	487	6	1.5	273	2	G87037	conserved hypothet
415	6	1.5	249	2	S72619	hypothetical prote	488	6	1.5	274	2	C84226	hypothetical prote
416	6	1.5	249	2	A80469	uroporphyrinogen-I	489	6	1.5	274	2	F83601	hypothetical prote
417	6	1.5	251	2	E96637	hypothetical prote	490	6	1.5	275	1	E36516	cis-1,2-dihydroben
418	6	1.5	252	2	C90422	hypothetical prote	491	6	1.5	275	2	D70896	probable deaA2 pro
419	6	1.5	252	2	B97479	ragA protein (Y096	492	6	1.5	275	2	T05595	hypothetical prote
420	6	1.5	252	2	A82697	two component resp	493	6	1.5	276	1	FLQ42C	hypothetical prote
421	6	1.5	252	2	T31439	probable cobyric a	494	6	1.5	276	2	S20690	31.6K hypothetical
422	6	1.5	252	2	AH3267	acetyltransferase	495	6	1.5	276	2	T29689	hypothetical prote
423	6	1.5	253	2	I54327	testicular protein	496	6	1.5	276	2	D70191	hypothetical prote
424	6	1.5	254	2	A12837	conserved hypothet	497	6	1.5	276	2	F83578	hypothetical prote
425	6	1.5	254	2	C97615	hypothetical prote	498	6	1.5	276	2	T44319	hypothetical prote
426	6	1.5	255	2	T39164	sorbitol utilizati	499	6	1.5	277	2	T40033	probable mitochond
427	6	1.5	255	2	B71490	hypothetical prote	500	6	1.5	278	2	D83080	hypothetical prote
428	6	1.5	255	2	A40571	testis-specific pr	501	6	1.5	279	2	T25587	hypothetical prote
429	6	1.5	256	2	S67040	probable membrane	502	6	1.5	280	1	A53419	2,3-dihydroxy-4-ph
430	6	1.5	256	2	T16805	hypothetical prote	503	6	1.5	281	2	F69400	2-deoxy-D-gluconat
431	6	1.5	257	2	T00780	myb-related protei	504	6	1.5	281	2	AC1490	conserved hypothet
432	6	1.5	257	2	E89124	protein K07C11.1 [	505	6	1.5	281	2	AP1129	conserved hypothet
433	6	1.5	257	2	B83292	hypothetical prote	506	6	1.5	282	2	JC5439	D-aspartate oxidas
434	6	1.5	257	2	F72500	hypothetical prote	507	6	1.5	282	2	T03227	hypothetical prote
435	6	1.5	258	1	E64479	cobalamin (5'-phos	508	6	1.5	282	2	B81264	probable DNA ligas
436	6	1.5	258	2	T11923	cytochrome-c oxida	509	6	1.5	283	2	S47845	rfaz protein - Esc
437	6	1.5	258	2	T24499	hypothetical prote	510	6	1.5	283	2	S42634	homeoprotein, IDX-
438	6	1.5	258	2	AH3079	ABC transporter, m	511	6	1.5	284	2	E97836	probable periplasm
439	6	1.5	259	2	S38988	cytochrome-c oxida	512	6	1.5	285	2	H96589	hypothetical prote
440	6	1.5	259	2	B81735	conserved hypothet	513	6	1.5	287	2	AD0612	anaerobic dimethyl
441	6	1.5	259	2	C70530	hypothetical prote	514	6	1.5	287	2	G90301	hypothetical prote
442	6	1.5	260	2	T32525	hypothetical prote	515	6	1.5	288	2	B87423	cytochrome c oxida
443	6	1.5	260	2	C82081	cell division prot	516	6	1.5	288	2	D86341	hypothetical prote
444	6	1.5	260	2	S00276	tall fiber protein	517	6	1.5	289	1	RLT2T	rRNA N-glycosidase
445	6	1.5	261	2	G90618	cytochrome c oxida	518	6	1.5	289	2	JC5606	karasurin C - Tric
446	6	1.5	261	2	G90620	cytochrome c oxida	519	6	1.5	289	2	T18919	hypothetical prote
447	6	1.5	261	2	T11081	cytochrome-c oxida	520	6	1.5	290	1	S70875	type IV prepilin p
448	6	1.5	261	2	T11199	cytochrome-c oxida	521	6	1.5	290	2	T16540	hypothetical prote
449	6	1.5	261	2	T11330	cytochrome-c oxida	522	6	1.5	290	2	S49588	isopentenyl-diphos
450	6	1.5	261	2	T11028	cytochrome-c oxida	523	6	1.5	291	2	F97300	probable membrane
451	6	1.5	261	2	T36011	hypothetical prote	524	6	1.5	292	2	E70314	probable heme A fa
452	6	1.5	261	2	T30158	hypothetical prote	525	6	1.5	292	2	G97325	transcription regu
453	6	1.5	261	2	T37571	hypothetical prote	526	6	1.5	293	2	S66356	mannose/glucose-bi
454	6	1.5	262	2	T11355	cytochrome-c oxida	527	6	1.5	293	2	A11998	tRNA pseudouridine
455	6	1.5	262	2	S00275	tall fiber protein	528	6	1.5	293	2	D96560	hypothetical prote
456	6	1.5	262	2	J70390	type II site-speci	529	6	1.5	294	2	AC1762	phosphate ABC tran
457	6	1.5	262	2	G82142	hypothetical prote	530	6	1.5	294	2	A11386	phosphate ABC tran
458	6	1.5	263	2	T38351	ribosomal protein	531	6	1.5	294	2	H82272	conserved hypothet
459	6	1.5	263	2	I41172	clpE protein - Esc	532	6	1.5	294	2	E69759	hypothetical prote
460	6	1.5	263	2	S64719	oncostatin M - mou	533	6	1.5	295	2	T33681	hypothetical prote
461	6	1.5	263	2	AC0522	probable secreted	534	6	1.5	295	2	T50204	conserved hypothet
462	6	1.5	265	2	B84468	hypothetical prote	535	6	1.5	295	2	T40854	polyaugar degradin
463	6	1.5	265	2	A98207	ABC transporter (p	536	6	1.5	296	2	A69856	hypothetical prote
464	6	1.5	265	2	G81377	probable transmem	537	6	1.5	296	2	D89795	hypothetical prote
465	6	1.5	265	2	AH2252	ATP-binding protei	538	6	1.5	297	2	T10410	envelope protein E
466	6	1.5	266	2	E71612	ribosomal protein	539	6	1.5	297	2	T28785	hypothetical prote
467	6	1.5	266	2	E72478	hypothetical prote	540	6	1.5	297	2	AB2879	hypothetical prote



541	6	1.5	299	1	XREC	ATP phosphoribosyl	614	6	1.5	324	2	A10434	probable sodium/ca
542	6	1.5	299	2	D90981	ATP phosphoribosyl	615	6	1.5	325	2	A82669	oxidative stress t
543	6	1.5	299	2	B85827	ATP phosphoribosyl	616	6	1.5	325	2	A19447	hypothetical prote
544	6	1.5	299	2	F96828	hypothetical prote	617	6	1.5	326	2	F69535	geranylgeranyl dip
545	6	1.5	299	2	B82600	conserved hypotet	618	6	1.5	326	2	H90938	hypothetical prote
546	6	1.5	300	2	A70819	hypothetical prote	619	6	1.5	326	2	D85787	hypothetical prote
548	6	1.5	301	2	B64164	hypothetical prote	620	6	1.5	326	2	A41862	hypothetical prote
549	6	1.5	301	2	S73023	hypothetical prote	621	6	1.5	326	2	T44594	C-S lyase (ORF326)
550	6	1.5	302	2	H87515	polysaccharide dea	622	6	1.5	326	2	AC3492	probable oxidoredu
551	6	1.5	302	2	A16455	3-hydroxybutyrate	623	6	1.5	327	2	H83659	acetoacetate-CoA 1
552	6	1.5	302	2	B49941	devr protein - Myx	624	6	1.5	328	1	C64580	serine/threonine-p
553	6	1.5	303	1	TVHUND	transforming prote	625	6	1.5	328	1	T26446	site-specific DNA-
554	6	1.5	303	2	E96560	hypothetical prote	626	6	1.5	328	2	T76446	hypothetical prote
555	6	1.5	304	2	G82237	ATP phosphoribosyl	627	6	1.5	328	2	D84563	oligopeptide trans
556	6	1.5	304	2	AD3597	ribose transport s	628	6	1.5	329	1	D84563	hypothetical prote
557	6	1.5	304	2	T34845	hypothetical prote	629	6	1.5	329	2	PN0448	zeta-crystallin /
558	6	1.5	304	2	E97143	probable hemin bin	630	6	1.5	329	2	AG1528	alcohol dehydrogen
559	6	1.5	304	2	E97143	hypothetical prote	631	6	1.5	329	2	AG1528	alcohol dehydrogen
560	6	1.5	305	2	G86022	probable periplasm	632	6	1.5	329	2	H64143	hypothetical prote
561	6	1.5	305	2	AH2863	conserved hypotet	633	6	1.5	330	1	H66840	competence protein
562	6	1.5	306	2	G95364	probable ABC trans	634	6	1.5	330	2	QOBE7	gene 62 protein -
563	6	1.5	306	2	AC0656	conserved hypotet	635	6	1.5	330	2	C86023	hypothetical prote
564	6	1.5	307	2	F90809	probable virulence	636	6	1.5	330	2	D82360	cell division prot
565	6	1.5	307	2	H70720	hypothetical prote	637	6	1.5	330	2	T31704	hypothetical prote
566	6	1.5	307	2	PN0470	coumarate-CoA liga	638	6	1.5	331	2	C84008	cation antiporter
567	6	1.5	307	2	B85669	probable virulence	639	6	1.5	331	2	AE2666	divalent cation tr
568	6	1.5	308	2	D83629	probable permease	640	6	1.5	333	2	F72617	probable dipeptide
569	6	1.5	308	2	D85879	hypothetical prote	641	6	1.5	334	2	F95902	probable ABC trans
570	6	1.5	308	2	B91035	hypothetical prote	642	6	1.5	335	2	B84432	hypothetical prote
571	6	1.5	308	2	T46294	hypothetical prote	643	6	1.5	335	2	AG0183	galactose-binding
572	6	1.5	311	2	S66468	catechol 1,2-dioxy	644	6	1.5	335	2	S54287	otm protein - Vib
573	6	1.5	311	2	T11017	MLC536_34 protein	645	6	1.5	335	2	G82224	probable vitamin B
574	6	1.5	311	2	H98278	hypothetical prote	646	6	1.5	335	2	T10208	hypothetical prote
575	6	1.5	311	2	AI3004	hypothetical prote	647	6	1.5	335	2	D98283	hypothetical prote
576	6	1.5	311	2	AI3004	hypothetical prote	648	6	1.5	336	2	AE3000	cell division prot
577	6	1.5	311	2	E82385	oxidoreductase, Gt	649	6	1.5	336	2	S74462	hypothetical prote
578	6	1.5	311	2	P97011	uncharacterized NA	650	6	1.5	336	2	G95003	membrane protein l
579	6	1.5	312	1	QOBE31L	probable membrane	651	6	1.5	336	2	A96997	glyceralddehyde-3-p
580	6	1.5	312	2	A42709	DNA modification m	652	6	1.5	337	1	DEBHG	glyceralddehyde-3-p
581	6	1.5	312	2	D90504	conserved hypotet	653	6	1.5	337	1	DEBHG	glyceralddehyde-3-p
582	6	1.5	312	2	D87475	rare lipoprotein A	654	6	1.5	337	1	DEBHG	glyceralddehyde-3-p
583	6	1.5	312	2	C71262	probable N-acetyl p	655	6	1.5	337	2	B83522	3-oxoacyl-l-acyl-ca
584	6	1.5	314	2	D97655	hemn-binding perit	656	6	1.5	337	2	C83059	conserved hypotet
585	6	1.5	314	2	S41428	cysteine proteinas	657	6	1.5	337	2	B95886	hypothetical prote
586	6	1.5	314	2	F81699	peptide ABC transp	658	6	1.5	338	1	AD3034	hypothetical prote
587	6	1.5	315	2	B83536	probable transcrip	659	6	1.5	338	1	DEIS3C	glyceralddehyde-3-p
588	6	1.5	315	2	B81937	probable transmemb	660	6	1.5	338	1	DEMDG	glyceralddehyde-3-p
589	6	1.5	315	2	B81168	transporter NMB070	661	6	1.5	338	2	AD4132	D-aspartate oxidas
590	6	1.5	316	1	G90406	protein F15E11.12	662	6	1.5	338	2	B90177	hypothetical prote
591	6	1.5	316	1	T08694	succinate dehydrog	663	6	1.5	338	2	S55657	hypothetical prote
592	6	1.5	316	2	B70768	hypothetical prote	664	6	1.5	338	2	D87182	glyceralddehyde-3-p
593	6	1.5	317	2	S19425	hypothetical prote	665	6	1.5	338	2	H86372	glutamate/glutamin
594	6	1.5	318	2	B91177	hypothetical prote	666	6	1.5	338	2	S77605	probable desal pro
595	6	1.5	318	2	G69596	probable hemin per	667	6	1.5	338	2	H70810	hypothetical prote
596	6	1.5	318	2	S48455	cephalosporin-C de	668	6	1.5	338	2	G71220	conserved hypotet
597	6	1.5	319	2	D69110	probable membrane	669	6	1.5	338	2	C89909	hypothetical prote
598	6	1.5	319	2	A70340	homoserine O-acety	670	6	1.5	339	2	H71265	hypothetical prote
599	6	1.5	320	2	A40660	conserved hypotet	671	6	1.5	340	1	UC5524	L-threonine 3-deny
600	6	1.5	320	2	E75255	outer membrane pro	672	6	1.5	340	1	A42987	N-acetyl-gamma-glu
601	6	1.5	321	2	T03431	hypothetical prote	673	6	1.5	340	1	D71154	hypothetical prote
602	6	1.5	321	2	AB3345	probable transport	674	6	1.5	340	2	F75007	suas related prote
603	6	1.5	321	2	F86571	hypothetical prote	675	6	1.5	340	2	A10392	thiamin biosynthes
604	6	1.5	321	2	B72051	methionyl tRNA for	676	6	1.5	340	2	B97876	conserved hypotet
605	6	1.5	321	2	F81615	methionyl tRNA for	677	6	1.5	341	1	DEJMG	glyceralddehyde-3-p
606	6	1.5	321	2	B69902	methionyl-CRNA for	678	6	1.5	341	2	UC5438	D-aspartate oxidas
607	6	1.5	321	2	S10006	probable sodium-de	679	6	1.5	341	2	D97082	branch-chain-aml
608	6	1.5	322	2	S66766	hypothetical prote	680	6	1.5	342	2	S57652	transcription fact
609	6	1.5	323	2	F75202	probable membrane	681	6	1.5	342	2	C87429	conserved hypotet
610	6	1.5	323	2	AD0994	dipeptide abc tran	682	6	1.5	343	2	F98064	2-dehydro-3-deoxy-
611	6	1.5	324	2	C70785	probable membrane	683	6	1.5	343	2	D70188	conserved hypotet
612	6	1.5	324	2	C87018	probable carboxydr	684	6	1.5	343	2	A64189	amino acid aminotr
613	6	1.5	324	2	D87075	probable DNA-bindi	685	6	1.5	343	2	T27798	hypothetical prote
							686	6	1.5	344	2	A11902	4-hydroxyphenylpyr

687 1.5 344 2 AH2828 HlyD family secret 760 6 1.5 364 2 A49781 cholesterol dehydr 364 2  
688 1.5 345 2 G87801 protein C10G11.4 [ 761 6 1.5 364 2 B83276 hypotheical prote 364 2  
689 1.5 346 2 A70649 probable dnf prote 762 6 1.5 364 2 AB1918 hypotheical prote 365 1  
690 1.5 346 2 A74019 hypotheical prote 763 6 1.5 365 1 B64228 hypotheical prote 366 2  
691 1.5 346 2 AC3292 metA/tolG/exbB pro 764 6 1.5 366 2 T21454 hypotheical prote 366 2  
692 1.5 347 2 A3815 transforming prote 765 6 1.5 366 2 B87446 hypotheical prote 366 2  
693 1.5 347 2 D83605 probable binding p 766 6 1.5 366 2 C84156 GTP-binding protei 366 2  
694 1.5 347 2 G83171 hypotheical prote 767 6 1.5 366 2 F72062 hypotheical prote 366 2  
695 1.5 347 2 T19685 hypotheical prote 768 6 1.5 366 2 A86561 CT449 hypotheical 367 2  
696 1.5 348 2 H98251 sugar ABC transpor 769 6 1.5 367 2 S28604 St12p protein - Ar 367 2  
697 1.5 349 2 A83531 DNA damage inducib 770 6 1.5 368 2 T01597 hypotheical prote 368 2  
698 1.5 349 2 T29843 hypotheical prote 771 6 1.5 369 2 H70850 hypotheical prote 369 2  
699 1.5 349 2 AH2858 glycosyltransferas 772 6 1.5 369 2 G82829 membrane fusion pr 369 2  
700 1.5 350 2 S60917 probable membrane 773 6 1.5 370 2 H70897 hypotheical prote 370 2  
701 1.5 350 2 T39555 hypotheical prote 774 6 1.5 370 2 E97622 gtn reductase (Y13 370 2  
702 1.5 350 2 T44206 hypotheical prote 775 6 1.5 370 2 AC2845 oxidoreductase Atu 370 2  
703 1.5 350 2 G33560 6R protein - human 776 6 1.5 370 2 E70354 probable GTP-bindi 370 2  
704 1.5 350 2 F90176 conserved hypothet 777 6 1.5 370 2 F87456 transcription regu 370 2  
705 1.5 350 2 D83563 hypotheical prote 778 6 1.5 370 2 S27344 hupK protein - Rhi 370 2  
706 1.5 350 2 B82230 conserved hypothet 779 6 1.5 370 2 AB3334 daunorubicin resis 370 2  
707 1.5 350 2 I78848 LIMK-2c - rat 780 6 1.5 371 2 T10635 3-hydroxyisobutyra 371 2  
708 1.5 351 2 T43111 hypotheical prote 781 6 1.5 371 2 I49698 alpha-1,3-galactos 371 2  
709 1.5 352 2 A12884 hypotheical prote 782 6 1.5 371 2 A99236 hypotheical 41.8K 371 2  
710 1.5 352 2 F37660 probable ABC trans 783 6 1.5 371 2 AB3050 conserved hypothet 371 2  
711 1.5 352 2 F69932 hypotheical prote 784 6 1.5 371 2 S75482 polyalialic acid tr 371 2  
712 1.5 352 2 T51330 DNA binding protei 785 6 1.5 371 2 H83285 probable two-compo 371 2  
713 1.5 353 2 H82644 NADP-alcohol dehyd 786 6 1.5 373 2 H84404 ferrichrome ABC tr 373 2  
714 1.5 353 2 F70046 iron permease homo 787 6 1.5 373 2 AD0030 conserved hypothet 373 2  
715 1.5 353 2 B3848 ferric ion ABC tra 788 6 1.5 373 2 S48496 saccharopine dehyd 373 2  
716 1.5 353 2 C84664 transferase, pepti 789 6 1.5 373 2 A69554 conserved hypothet 373 2  
717 1.5 353 2 G71852 udp-n-acetylglucos 790 6 1.5 374 2 B64418 N2,N2-dimethylguan 374 2  
718 1.5 353 2 AF1759 aldose 1-epimerase 791 6 1.5 374 2 E95861 hypotheical ABC t 374 2  
719 1.5 353 2 A10220 probable sugar tra 792 6 1.5 374 2 AE2227 transposase alx337 374 2  
720 1.5 353 2 D75328 conserved hypothet 793 6 1.5 375 1 T44848 call division prot 375 1  
721 1.5 353 2 G87392 conserved hypothet 794 6 1.5 375 2 AC0337 probable D-isomer 375 2  
722 1.5 353 2 S78304 protoporphylin IX 795 6 1.5 375 2 H84179 cell division prot 375 2  
723 1.5 353 2 S37138 protoporphylin IX 796 6 1.5 375 2 A40112 MHC-region RD-repe 375 2  
724 1.5 353 2 S52520 hypotheical prote 797 6 1.5 376 2 C82310 sulfate ABC tranep 376 2  
725 1.5 354 2 A48931 transmembrane glyc 798 6 1.5 376 2 H82308 oxalacetate decar 376 2  
726 1.5 354 2 E75582 hypotheical prote 799 6 1.5 377 2 F97001 endoglucanase fami 377 2  
727 1.5 354 2 C96600 protein F14J16.16 800 6 1.5 377 2 T49604 hypotheical prote 377 2  
728 1.5 354 2 G97640 hypotheical prote 801 6 1.5 379 2 A90337 proteinase, probab 379 2  
729 1.5 355 2 T13862 NADH2 dehydrogenas 802 6 1.5 380 2 T35953 conserved hypothet 380 2  
730 1.5 355 2 S06939 F14P1.7 protein - 803 6 1.5 380 2 S22415 membrane protein T 380 2  
731 1.5 355 2 F86328 hypotheical prote 804 6 1.5 380 2 S36789 transcripion nega 380 2  
732 1.5 356 2 A86590 hypotheical prote 805 6 1.5 381 2 T02589 hypotheical prote 381 2  
733 1.5 356 2 D72033 hypotheical prote 806 6 1.5 382 2 S32148 exo-alpha-glialdas 382 2  
734 1.5 356 2 A88230 hypotheical prote 807 6 1.5 383 2 H83626 probable RND efflu 383 2  
735 1.5 356 2 A83036 hypotheical prote 808 6 1.5 383 2 F97606 hypotheical prote 383 2  
736 1.5 357 2 S18828 band 3 anion trans 809 6 1.5 384 2 C81741 nifs protein, prob 384 2  
737 1.5 357 2 H83094 UDP-N-acetylglucos 810 6 1.5 384 2 AH3206 HlyD family secret 384 2  
738 1.5 357 2 H98819 conserved hypothet 811 6 1.5 385 2 B88951 protein C38C3.3 [i 385 2  
739 1.5 357 2 E90815 hypotheical prote 812 6 1.5 386 1 SYECGG O-succinylhomoseri 386 1  
740 1.5 357 2 A85675 hypotheical prote 813 6 1.5 386 2 D91X37 cystathionine gamm 386 2  
741 1.5 357 2 H64855 probable membrane 814 6 1.5 386 2 H86084 cystathionine gamm 386 2  
742 1.5 358 2 A93332 malate dehydrogena 815 6 1.5 386 2 A10937 cystathionine gamm 386 2  
743 1.5 359 2 P00468 threonine ammonia- 816 6 1.5 386 2 E70676 probable nonspecif 386 2  
744 1.5 359 2 H64547 peptide methionine 817 6 1.5 387 2 A82014 capsule polysaccha 387 2  
745 1.5 359 2 E71960 probable peptide m 818 6 1.5 387 2 G83873 hypotheical prote 387 2  
746 1.5 359 2 T21247 hypotheical prote 819 6 1.5 387 2 S61845 probable membrane 387 2  
747 1.5 360 2 S49330 serine/threonine k 820 6 1.5 390 2 B72639 hypotheical prote 390 2  
748 1.5 360 2 G82967 phosphoribosylamin 821 6 1.5 390 2 C81318 probable galactosy 390 2  
749 1.5 360 2 B98217 peptide chain rele 822 6 1.5 391 2 S23341 paired box transcr 391 2  
750 1.5 360 2 AH3069 peptide chain rele 823 6 1.5 391 2 A44063 paired box transcr 391 2  
751 1.5 360 2 C82445 hypotheical prote 824 6 1.5 391 2 S15220 capsule polysaccha 391 2  
752 1.5 360 2 T18529 protein phosphatog 825 6 1.5 391 2 G90266 na(+)/H(+) antipor 391 2  
753 1.5 361 2 A40023 paired box homolog 826 6 1.5 392 2 S70197 UNFGlucose 6-dehyd 392 2  
754 1.5 361 2 B56940 integral membrane 827 6 1.5 392 2 G75452 probable glutamate 392 2  
755 1.5 362 2 C97448 hypotheical prote 828 6 1.5 392 2 B69321 cell division prot 392 2  
756 1.5 363 1 G69869 xaa-Pro dipeptidas 829 6 1.5 392 2 T38450 probable GTP-bindi 392 2  
757 1.5 363 2 T06726 cysteine proteinas 830 6 1.5 393 2 F72068 dihydrolipoamide s 393 2  
758 1.5 363 2 A56940 integral membrane 831 6 1.5 393 2 C86556 dihydrolipoamide s 393 2  
759 1.5 363 2 F97635 alpha-D-mannose-al 832 6 1.5 393 2 C86556 dihydrolipoamide s 393 2

833	6	1.5	393	2	B72455	threonine synthase
834	6	1.5	393	2	A49008	paired box transcr
835	6	1.5	393	2	T23602	hypothetical prote
836	6	1.5	393	2	B84425	probable protein t
837	6	1.5	393	2	T46907	StiIIp protein [imp
838	6	1.5	394	2	E71136	threonine synthase
839	6	1.5	394	2	G75080	threonine synthase
840	6	1.5	394	2	A34417	alpha-1,3-mannosyl
841	6	1.5	394	2	A47223	pancothenate metab
842	6	1.5	394	2	S45881	probable purine nu
843	6	1.5	394	2	H65010	hypothetical prote
844	6	1.5	394	2	S72814	hypothetical prote
845	6	1.5	396	2	C69414	threonine synthase
846	6	1.5	396	2	AH3175	serine-type D-Ala-
847	6	1.5	396	2	B32808	flagellin chain B
848	6	1.5	396	2	T46901	hypothetical prote
849	6	1.5	396	2	F86183	hypothetical prote
850	6	1.5	397	2	G70796	hypothetical prote
851	6	1.5	397	2	D70311	probable phosphogl
852	6	1.5	397	2	T20013	hypothetical prote
853	6	1.5	398	2	B75254	acetyl kinase - D
854	6	1.5	398	2	A37768	streptococcal pyro
855	6	1.5	398	2	T53340	paired box transcr
856	6	1.5	398	2	T47967	hypothetical prote
857	6	1.5	398	2	D81397	probable periplasm
858	6	1.5	399	2	A83354	hypothetical prote
859	6	1.5	399	2	A53856	aryl-acylamidase (
860	6	1.5	399	2	E72453	hypothetical prote
861	6	1.5	399	2	A11114	surface protein (p
862	6	1.5	402	2	T24184	hypothetical prote
863	6	1.5	402	2	T21278	hypothetical prote
864	6	1.5	402	2	AG1839	hypothetical prote
865	6	1.5	403	2	B72778	probable processin
866	6	1.5	403	2	B36151	tryptophan synthas
867	6	1.5	403	2	F64377	cell division proc
868	6	1.5	403	2	F90283	hypothetical prote
869	6	1.5	403	2	S07825	hypothetical prote
870	6	1.5	404	2	G83322	hypothetical prote
871	6	1.5	404	2	S38127	hypothetical prote
872	6	1.5	404	2	T35421	probable regulator
873	6	1.5	405	2	H64482	threonine synthase
874	6	1.5	405	2	D97563	hypothetical prote
875	6	1.5	405	2	AD2784	molybdopterin bios
876	6	1.5	405	2	H71692	hypothetical prote
877	6	1.5	405	2	T15601	hypothetical prote
878	6	1.5	406	2	T36196	probable acyl carr
879	6	1.5	407	2	T49708	probable 3-methyl-
880	6	1.5	407	2	G84635	probable protein k
881	6	1.5	407	2	T37104	hypothetical glycosyl
882	6	1.5	407	2	T19176	hypothetical prote
883	6	1.5	408	2	B87436	conserved hypotet
884	6	1.5	408	2	H95416	probable ROK famli
885	6	1.5	409	2	B89354	hypothetical prote
886	6	1.5	409	2	T16781	hypothetical prote
887	6	1.5	410	2	S70119	hypothetical prote
888	6	1.5	410	2	P96683	hypothetical prote
889	6	1.5	410	2	C69197	ATP-dependent 26S
890	6	1.5	411	2	S58094	hypothetical MW do
891	6	1.5	411	2	B83043	hypothetical prote
892	6	1.5	411	2	F97760	hypothetical prote
893	6	1.5	412	2	AF2320	hypothetical prote
894	6	1.5	412	2	S28423	isocitrate dehydro
895	6	1.5	412	2	T10671	protein kinase hom
896	6	1.5	412	2	S72517	alcohol dehydrogen
897	6	1.5	413	2	S52251	paired box transcr
898	6	1.5	413	2	E70661	probable PB protei
899	6	1.5	413	2	D81957	amino-acid transpo
900	6	1.5	413	2	AG0632	glucose-1-phosphat
901	6	1.5	414	2	B64609	RNA polymerase sig
902	6	1.5	414	2	F71905	RNA polymerase sig
903	6	1.5	414	2	B90186	phosphoenolpyruvate
904	6	1.5	415	1	A31133	diaminopimelate de
905	6	1.5	415	2	S65065	isocitrate dehydro
906	6	1.5	415	2	F82986	diaminopimelate de
907	6	1.5	415	2	A60086	paired box transcr
908	6	1.5	415	2	D85061	probable adenosine
909	6	1.5	415	2	B85436	hypothetical prote
910	6	1.5	416	2	F90076	hypothetical prote
911	6	1.5	416	2	S47013	isocitrate dehydro
912	6	1.5	416	2	T10458	diaminopimelate de
913	6	1.5	416	2	H81376	anthranilate synth
914	6	1.5	416	2	D71277	hypothetical prote
915	6	1.5	416	2	T28756	hypothetical prote
916	6	1.5	416	2	T16205	hypothetical prote
917	6	1.5	417	2	B83806	homoserine dehydro
918	6	1.5	417	2	H83708	hypothetical prote
919	6	1.5	419	1	S70617	homeotic protein g
920	6	1.5	420	2	C84201	multidrug resistan
921	6	1.5	420	2	AC0958	probable membrane
922	6	1.5	421	2	B69467	hypothetical prote
923	6	1.5	421	2	F72213	hypothetical prote
924	6	1.5	422	2	T33374	hypothetical prote
925	6	1.5	423	2	C70582	probable PPS prote
926	6	1.5	425	1	A26431	nerve growth facto
927	6	1.5	425	2	A38153	paired box protei
928	6	1.5	425	2	A64625	pancothenate metab
929	6	1.5	425	2	T50355	hypothetical prote
930	6	1.5	426	2	H69127	histidinol dehydro
931	6	1.5	426	2	T48424	hypothetical prote
932	6	1.5	427	1	G0HUN	nerve growth facto
933	6	1.5	427	2	T14424	S-luciferase
934	6	1.5	428	2	AG1676	uracil permease ho
935	6	1.5	428	2	T27544	zinc resistance pr
936	6	1.5	428	2	AH3536	dihydroipoamide S
937	6	1.5	429	2	S49595	probable serine pr
938	6	1.5	429	2	S48172	variable surface g
939	6	1.5	429	2	AH0630	4-hydroxyphenylase
940	6	1.5	429	2	AB2456	glycosyltransferas
941	6	1.5	430	2	H83178	conserved hypotet
942	6	1.5	433	1	VHVUPH	nucleocapsid prote
943	6	1.5	433	2	T09619	isocitrate dehydro
944	6	1.5	433	2	T14329	dermal glycoprotei
945	6	1.5	433	2	D82280	oxaloacetate decar
946	6	1.5	433	2	S69999	sterigmatocystin g
947	6	1.5	434	2	A57986	membrane-bound ura
948	6	1.5	435	2	S30792	Npl6 protein - yea
949	6	1.5	435	2	AF0630	4-Hydroxyphenylase
950	6	1.5	436	2	T00908	hypothetical prote
951	6	1.5	437	1	Z4BPIK	gene IV protein -
952	6	1.5	437	2	A75194	glycogen synthase
953	6	1.5	437	2	AF2330	preprotein translo
954	6	1.5	438	2	T07402	probable isocitrat
955	6	1.5	440	2	D83089	histidinol dehydro
956	6	1.5	440	2	F86385	probable lysine an
957	6	1.5	441	2	T42946	hypothetical prote
958	6	1.5	442	1	B43362	protein-tyrosine k
959	6	1.5	443	1	E64667	multidrug-efflux t
960	6	1.5	443	2	F71848	probable transpore
961	6	1.5	444	2	C83336	probable cytochrom
962	6	1.5	444	2	H86224	hypothetical prote
963	6	1.5	444	2	S75455	Na+-ATPase chain J
964	6	1.5	445	2	AD2031	hypothetical prote
965	6	1.5	445	2	AB2022	hypothetical prote
966	6	1.5	446	2	T34782	probable signal pe
967	6	1.5	447	2	AH0536	conserved hypotet
968	6	1.5	449	2	S57084	actin-related prot
969	6	1.5	449	2	G64349	phosphomannomutase
970	6	1.5	449	2	F83627	hypothetical prote
971	6	1.5	449	2	T39410	hypothetical prote
972	6	1.5	449	2	H88022	protein T27A1.5 (I
973	6	1.5	449	2	T08865	hypothetical prote
974	6	1.5	450	2	F90210	hypothetical prote
975	6	1.5	450	2	G96749	hypothetical prote
976	6	1.5	450	2	A54429	paired box transcr
977	6	1.5	450	2	T37628	glycerol dehydroge
978	6	1.5	451	1	JE0240	lim kinase (EC 2.7

```
979      6      1.5      451      2      S33612
980      6      1.5      451      2      S15236
981      6      1.5      451      2      E96763
982      6      1.5      452      2      G95347
983      6      1.5      454      2      A82353
984      6      1.5      454      2      A82994
985      6      1.5      455      2      G70089
986      6      1.5      455      2      S33033
987      6      1.5      456      1      QXFF73
988      6      1.5      456      2      A64052
989      6      1.5      456      2      H86406
990      6      1.5      456      2      D71537
991      6      1.5      456      2      F81692
992      6      1.5      456      2      AG3380
993      6      1.5      457      2      S18239
994      6      1.5      457      2      T09932
995      6      1.5      457      2      A56925
996      6      1.5      458      2      E82444
997      6      1.5      458      2      F84746
998      6      1.5      458      2      F97296
999      6      1.5      458      2      H75603
1000     6      1.5      459      2      S52250

isocitrate dehydro
glutathione-disulf
unknown protein F2
Protein [imported]
glutathione-disulf
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical 53K p
glutathione-disulf
probable serine ca
probable replicati
chromosomal replic
outer membrane pro
glycerol-3-phospha
probable phosphodi
paired box transcr
NAD(P) transhydrog
probable serine ca
conserved hypothet
paired box transcr

ALIGNMENTS

RESULT 1
S01042
aerobactin receptor precursor iutA [validated] - Escherichia coli plasmid Colv-K30
N:Alternate names: cloacin receptor
C:Species: Escherichia coli
C>Date: 31-Dec-1998 #sequence_revision 31-Dec-1998 #text_change 26-May-2000
C:Accession: S01042
R:Krone, W.J.A.; Stegehuis, F.; Koningsstein, G.; van Doorn, C.; Roosendaal, B.; de Graaf
FEMS Microbiol. Lett. 26, 153-161, 1998
A:Title: Characterization of the pColv-K30 encoded cloacin DF13/aerobactin outer membran
e sequence and primary structure.
A:Reference number: S01042
A:Accession: S01042
A:Molecule type: DNA
A:Residues: 1-725 <KRO>
A:Cross-references: EMBL:X05974
C:Genetics:
A:Gene: iutA
A:Genome: plasmid Colv-K30
C:Function:
A:Description: functions as outer membrane receptor for ferric aerobactin [validated, MU
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-725/Product: cloacin receptor #status predicted <MAT>

Query Match      2.5%; Score 10; DB 2; Length 725;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 154 TGGLINIVTK 163
Db 144 TGGLINIVTK 153

RESULT 2
AB0122
probable ferric siderophore receptor iutA [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AB0122
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AB0122

Query Match      2.5%; Score 10; DB 2; Length 725;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 154 TGGLINIVTK 163
Db 144 TGGLINIVTK 153

RESULT 3
T44483
receptor-like protein iutA [imported] - Shigella flexneri
C:Species: Shigella flexneri
C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T44483
R:Moss, J.E.; Cardozo, T.J.; Zychlinsky, A.; Groisman, E.A.
Mol. Microbiol. 33, 74-83, 1999
A:Title: The selc-associated SHI-2 pathogenicity island of Shigella flexneri.
A:Reference number: 222779; MUID:99340540; PMID:10411725
A:Accession: T44483
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-732 <MOS>
A:Cross-references: EMBL:AF141323; NID:G5532445; PIDN:AA044750.1; PID:G5532466
A:Experimental source: strain M907; serotype 5a
C:Genetics:
A:Gene: iutA

Query Match      2.5%; Score 10; DB 2; Length 732;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 154 TGGLINIVTK 163
Db 143 TGGLINIVTK 152

RESULT 4
F83476
probable siderophore receptor PA1365 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83476
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: F83476
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-813 <STO>
A:Cross-references: GB:AE004565; GB:AE004091; NID:G9947294; PIDN:AAG04754.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1365

Query Match      2.2%; Score 9; DB 2; Length 813;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 148 IYGSATGCG 156
Db 229 IYGSATGCG 237
```

```
979      6      1.5      451      2      S33612
980      6      1.5      451      2      S15236
981      6      1.5      451      2      E96763
982      6      1.5      452      2      G95347
983      6      1.5      454      2      A82353
984      6      1.5      454      2      A82994
985      6      1.5      455      2      G70089
986      6      1.5      455      2      S33033
987      6      1.5      456      1      QXFF73
988      6      1.5      456      2      A64052
989      6      1.5      456      2      H86406
990      6      1.5      456      2      D71537
991      6      1.5      456      2      F81692
992      6      1.5      456      2      AG3380
993      6      1.5      457      2      S18239
994      6      1.5      457      2      T09932
995      6      1.5      457      2      A56925
996      6      1.5      458      2      E82444
997      6      1.5      458      2      F84746
998      6      1.5      458      2      F97296
999      6      1.5      458      2      H75603
1000     6      1.5      459      2      S52250

isocitrate dehydro
glutathione-disulf
unknown protein F2
Protein [imported]
glutathione-disulf
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical 53K p
glutathione-disulf
probable serine ca
probable replicati
chromosomal replic
outer membrane pro
glycerol-3-phospha
probable phosphodi
paired box transcr
NAD(P) transhydrog
probable serine ca
conserved hypothet
paired box transcr

ALIGNMENTS

RESULT 1
S01042
aerobactin receptor precursor iutA [validated] - Escherichia coli plasmid Colv-K30
N:Alternate names: cloacin receptor
C:Species: Escherichia coli
C>Date: 31-Dec-1998 #sequence_revision 31-Dec-1998 #text_change 26-May-2000
C:Accession: S01042
R:Krone, W.J.A.; Stegehuis, F.; Koningsstein, G.; van Doorn, C.; Roosendaal, B.; de Graaf
FEMS Microbiol. Lett. 26, 153-161, 1998
A:Title: Characterization of the pColv-K30 encoded cloacin DF13/aerobactin outer membran
e sequence and primary structure.
A:Reference number: S01042
A:Accession: S01042
A:Molecule type: DNA
A:Residues: 1-725 <KRO>
A:Cross-references: EMBL:X05974
C:Genetics:
A:Gene: iutA
A:Genome: plasmid Colv-K30
C:Function:
A:Description: functions as outer membrane receptor for ferric aerobactin [validated, MU
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-725/Product: cloacin receptor #status predicted <MAT>

Query Match      2.5%; Score 10; DB 2; Length 725;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 154 TGGLINIVTK 163
Db 144 TGGLINIVTK 153

RESULT 2
AB0122
probable ferric siderophore receptor iutA [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001
C:Accession: AB0122
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AB0122
```

RESULT 5  
T29061  
Probable integral membrane protein - Streptomyces coelicolor  
C/Species: Streptomyces coelicolor  
C/Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 11-May-2000  
C/Accession: T29061  
R/Redenbach, M.; Kleser, H.M.; Denapate, D.; Eichner, A.; Cullum, J.; Kinashi, H.; Hopw  
Mol. Microbiol. 21, 77-96, 1996  
A/Title: A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb S  
A/Reference number: Z20556; MUID:97000351; PMID:8843436  
A/Accession: T29061  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-334 <RED>  
A/Cross-references: EMBL:AL031124; NID:el312893; PID:el312905; PIDN:CAA1979.1  
A/Note: SCIC2.12c

Query Match 2.0%; Score 8; DB 2; Length 334;  
Best Local Similarity 100.0%; Pred. No. 7.5;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 LGVSSGTT 97  
|||||  
Db 87 LGVSSGTT 94

RESULT 6  
B25866  
Short neurotoxin II - yellow-lipped sea krait  
C/Species: Laticauda colubrina (yellow-lipped sea krait)  
C/Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 07-Feb-1997  
C/Accession: B25866  
R/Tamiya, N.; cited by Endo, T.; Nakanishi, M.; Furukawa, S.; Joubert, F.J.; Tamiya, N.;  
Biochemistry 25, 395-404, 1986  
A/Title: Stopped-flow fluorescence studies on binding kinetics of neurotoxins with acety  
A/Reference number: A90508; MUID:86159724; PMID:3955004  
A/Accession: B25866  
A/Molecule type: protein  
A/Residues: 1-62 <TRM>  
C/Superfamily: snake toxin  
C/Keywords: neurotoxin

Query Match 1.8%; Score 7; DB 2; Length 62;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 301 QPKTKS 307  
|||||  
Db 10 QPKTKS 16

RESULT 7  
D25866  
Short neurotoxin d - yellow-lipped sea krait  
C/Species: Laticauda colubrina (yellow-lipped sea krait)  
C/Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 07-Feb-1997  
C/Accession: D25866  
R/Tamiya, N.; cited by Endo, T.; Nakanishi, M.; Furukawa, S.; Joubert, F.J.; Tamiya, N.;  
Biochemistry 25, 395-404, 1986  
A/Title: Stopped-flow fluorescence studies on binding kinetics of neurotoxins with acety  
A/Reference number: A90508; MUID:86159724; PMID:3955004  
A/Accession: D25866  
A/Molecule type: protein  
A/Residues: 1-62 <TRM>  
C/Superfamily: snake toxin  
C/Keywords: neurotoxin

Query Match 1.8%; Score 7; DB 2; Length 62;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 301 QPKTKS 307  
|||||  
Db 10 QPKTKS 16

RESULT 8  
C25866  
Short neurotoxin c - yellow-lipped sea krait  
C/Species: Laticauda colubrina (yellow-lipped sea krait)  
C/Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 07-Feb-1997  
C/Accession: C25866  
R/Tamiya, N.; cited by Endo, T.; Nakanishi, M.; Furukawa, S.; Joubert, F.J.; Tamiya, N.;  
Biochemistry 25, 395-404, 1986  
A/Title: Stopped-flow fluorescence studies on binding kinetics of neurotoxins with acety  
A/Reference number: A90508; MUID:86159724; PMID:3955004  
A/Accession: C25866  
A/Molecule type: protein  
A/Residues: 1-62 <TRM>  
C/Superfamily: snake toxin  
C/Keywords: neurotoxin

Query Match 1.8%; Score 7; DB 2; Length 62;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 301 QPKTKS 307  
|||||  
Db 10 QPKTKS 16

RESULT 9  
AG0748  
Hypothetical protein SRY2148 [imported] - Salmonella enterica subsp. enterica serovar Typh  
C/Species: Salmonella enterica subsp. enterica serovar Typh  
A/Note: this species has also been called Salmonella typh  
C/Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 09-Nov-2001  
C/Accession: AG0748  
R/Parhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Main, J.; Churcher,  
th, T.; Connerton, P.; Cronin, A.; Davies, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
S.; Moule, S.; O'Garra, P.  
Nature 413, 848-852, 2001  
A/Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
A/Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
A/Reference number: AB0502; PMID:11677608  
A/Accession: AG0748  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-63 <PAR>  
A/Cross-references: GB:AL513382; PIDN:CAD05690.1; PID:g16503184; GSPDB:GN00176  
C/Genetics:  
A/Gene: SRY2148

Query Match 1.8%; Score 7; DB 2; Length 63;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 SLPPLSV 17  
|||||  
Db 43 SLPPLSV 49

RESULT 10  
G86631  
Hypothetical protein yafU [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
C/Species: Lactococcus lactis subsp. lactis  
C/Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 03-Aug-2001  
C/Accession: G86631  
R/Bolotin, A.; Minkler, P.; Mauger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; Ehrlich  
Genome Res. 11, 731-753, 2001  
A/Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis 88  
A/Reference number: AB6625; MUID:21235186; PMID:11337471  
A/Accession: G86631  
A/Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-87 <STO>  
A:Cross-references: GB:AE005176; PID:g12722896; PIDN:AAK04153.1; GSPDB:GN00146  
A:Experimental source: strain IL1403  
C:Genetics:  
A:Gene: yafu

Query Match 1.8%; Score 7; DB 2; Length 87;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

Qy 91 GVSSGTT 97  
|||||  
Db 43 GVSSGTT 49

RESULT 11  
H86382  
hypotheical protein F4F7.16 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
C:Accession: H86382  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Dewar, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: H86382  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-131 <STO>  
A:Cross-references: GB:AE005172; NID:g11067284; PIDN:AAG28812.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1

Query Match 1.8%; Score 7; DB 2; Length 131;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 279 RLAVLFG 285  
|||||  
Db 116 RLAVLFG 122

RESULT 12  
T18204  
mnxB protein - Bacillus sp.  
C:Species: Bacillus sp.  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T18204  
R:van Waasbergen, L.G.; Hildebrand, M.; Tebo, B.M.  
J. Bacteriol. 178, 3517-3530, 1996  
A:Title: Identification and characterization of a gene cluster involved in manganese oxida  
A:Reference number: Z18825; MUID:96256605; PMID:8655549  
A:Accession: T18204  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-154 <VAN>  
A:Cross-references: EMBL:U31081; NID:g942616; PID:g942618; PIDN:AAB06484.1  
C:Genetics:  
A:Gene: mnxB  
A:Function:  
A:Description: may be involved in manganese oxidation by spores

Query Match 1.8%; Score 7; DB 2; Length 154;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 SGATGGL 157  
|||||  
Db 95 SGATGGL 101

## RESULT 13

A46189  
vasotocin - Pacific hagfish  
C:Species: Eptatretus stouti (Pacific hagfish)  
C>Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Mar-1997  
C:Accession: A46189  
R:Heierhorst, J.; Lederis, K.; Richter, D.  
Proc. Natl. Acad. Sci. U.S.A. 89, 6798-6802, 1992  
A:Title: Presence of a member of the Tc1-like transposon family from nematodes and Drosoph  
A:Reference number: A46189; MUID:92357719; PMID:1379721  
A:Accession: A46189  
A>Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-161 <HEI>  
A:Experimental source: hypothalamus  
A>Note: sequence extracted from NCBI backbone (NCBIN:110199, NCBIP:110200)  
C:Superfamily: oxytocin-neurophysin  
C:Keywords: neuropeptide

Query Match 1.8%; Score 7; DB 2; Length 161;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 SGATGGL 157  
|||||  
Db 135 SGATGGL 141

## RESULT 14

B84348  
hypotheical protein Vngl974h [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: B84348  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Laaky, S.  
; Leithauer, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon  
Jung, K.H.; Alam, M.; Fritas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
A:Title: Genome sequence of Halobacterium species NRC-1.  
A:Reference number: A84160; MUID:20504483; PMID:11016950  
A:Accession: B84348  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-173 <STO>  
A:Cross-references: GB:AE004437; NID:g10581409; PIDN:AAG20150.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: VNG1974H

Query Match 1.8%; Score 7; DB 2; Length 173;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 216 GGAFDAN 222  
|||||  
Db 118 GGAFDAN 124

## RESULT 15

A87268  
grpE protein [imported] - Caulobacter crescentus  
C:Species: Caulobacter crescentus  
C>Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 20-Apr-2001  
C:Accession: A87268  
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J  
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
A>Title: Complete Genome Sequence of *Caulobacter crescentus*.  
A:Reference number: A87249; MUID:21173698; PMID:11259647  
A:Accession: A87268  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-198 <STO>  
A:Cross-References: GB:AE005673; NID:g13421267; PIDN:PAK22141.1; GSPDB:GN00148  
C:Genetics:  
A:Gene: CC0154

Query Match 1.8%; Score 7; DB 2; Length 198;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 TGGAFDA 221  
DB 190 TGGAFDA 196

## RESULT 16

hypothetical protein PA683 [imported] - *Pseudomonas aeruginosa* (strain PA01)

C:Species: *Pseudomonas aeruginosa*  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000

C:Accession: H83059

R:Stover, C.K.; Pham, X.O.; Ewitt, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Lardig, K.; Lim,  
., Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000

A>Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho

A:Reference number: A82950; MUID:20437337; PMID:10984043  
A:Accession: H83059

A>Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-200 <STO>  
A:Experimental source: strain PA01

C:Genetics:  
A:Gene: PA683

Query Match 1.8%; Score 7; DB 2; Length 200;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 SLNAIKG 295  
DB 181 SLNAIKG 187

## RESULT 17

hypothetical protein all5245 [imported] - *Nostoc* sp. (strain PCC 7120)

C:Species: *Nostoc* sp.

A>Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002

C:Accession: AE2461

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kunitz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi,  
M.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata,  
S.  
DNA Res. 8, 205-213, 2001

A>Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Ana*

A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AE2461

A>Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-223 <KOR>  
A:Cross-References: GB:BA000019; PIDN:BAB76944.1; PID:g17134384; GSPDB:GN00179

A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: all5245

Query Match 1.8%; Score 7; DB 2; Length 223;  
Best Local Similarity 100.0%; Pred. No. 56;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 390 PNLNPKK 396  
DB 30 PNLNPKK 36

## RESULT 18

hypothetical protein SCH24.22c - *Streptomyces coelicolor*

C:Species: *Streptomyces coelicolor*  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999

C:Accession: T36587

R:Olliver, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, May 1999

A:Reference number: 221575  
A:Accession: T36587

A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA

A:Residues: 1-225 <OLI>  
A:Cross-References: EMBL:AL049826; PIDN:CAB42728.1; GSPDB:GN00070; SCOEDB:SCH24.22c

A:Experimental source: strain A3 (2)  
C:Genetics:  
A:Gene: SCOEDB:SCH24.22c

Query Match 1.8%; Score 7; DB 2; Length 225;  
Best Local Similarity 100.0%; Pred. No. 57;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 SGATGGL 157  
DB 202 SGATGGL 208

## RESULT 19

conserved hypothetical protein XF0590 [imported] - *Xylella fastidiosa* (strain 9A5C)

C:Species: *Xylella fastidiosa*

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: H82789

R:anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000  
A>Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.

A:Reference number: A82515; MUID:20365717; PMID:10910347  
A>Note: for a complete list of authors see reference number A59328 below

A:Accession: H82789

A>Status: preliminary  
A:Molecule type: DNA

A:Residues: 1-233 <SIM>  
A:Cross-References: GB:AE003904; GB:AE003849; NID:g9105433; PIDN:AAF83400.1; GSPDB:GN001

A:Experimental source: strain 9A5C  
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; B  
riones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carro, D.M.; Carrer, F  
de-Nevo, E.; Docena, C.; El-Dorri, H.; Faciniani, A.P.; Ferreira, A.J.S.  
submitted to GenBank, June 2000

A:Author: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
J.D.; Unquifera, M.L.; Kemper, E.L.; Kitejima, J.P.; Kiege, J.B.; Kuramae, E.B.; Laig  
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, F  
A:Authors: Martins, E.M.F.; Matsumura, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.  
, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeira, D.J  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasat  
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
M.; Tenhake, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.;

A:Reference number: A59328  
A:Contents: annotation  
C:Genetics:  
A:Gene: XF0590

Query Match 1.8%; Score 7; DB 2; Length 233;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 190 QVGQSYA 196

Db 67 QVGQSV 73

RESULT 20

DSHUEC

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 20-Apr-2000

C:Accession: A28301; A54656; I55405

R: Hjalmarsson, K.; Marklund, S.L.; Engstroem, A.; Edlund, T.

Proc. Natl. Acad. Sci. U.S.A. 84, 6340-6344, 1987

A:Title: Isolation and sequence of complementary DNA encoding human extracellular superoxide dismutase

A:Reference number: A28301; MUID:87317647; PMID:3476950

A:Accession: A28301

A:Molecule type: mRNA

A:Residues: 1-240 <HJA>

A:Cross-references: GB:J02947; NID:G338283; PIDN:AAA66000.1; PID:G338284

R: Polz, R.J.; Crapo, J.D.

Genomics 22, 162-171, 1994

A:Title: Extracellular superoxide dismutase (SOD3): tissue-specific expression, genomic

A:Reference number: A54656; MUID:95048365; PMID:7959763

A:Accession: A54656

A:Molecule type: DNA

A:Residues: 1-39 <POL>

A:Cross-references: GB:U10116

R: Sandstrom, J.; Nilsson, P.; Karlsson, K.; Marklund, S.L.

J. Biol. Chem. 269, 19163-19166, 1994

A:Title: 10-fold increase in human plasma extracellular superoxide dismutase content

A:Reference number: I55405; MUID:94308185; PMID:8034674

A:Accession: I55405

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 195-230, 'G', 232-240 <RES>

A:Cross-references: GB:S71544; NID:G565135; PIDN:AAB31336.1; PID:G565136

A:Note: mutant sequence

C:Genetics:

A:Gene: GDB:SOD3

A:Cross-references: GDB:125291; OMIM:185490

A:Map position: 4p16.3-4q21

A:Introns: #status absent

A:Note: the two introns occur before the initiator codon

C:Complex: homotetramer

C:Function:

A:Description: catalyzes the dismutation of 2 molecules of peroxide radical to dioxygen

C:Superfamily: superoxide dismutase (Cu-Zn)

C:Keywords: copper; glycoprotein; metalloprotein; oxidoreductase; zinc

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-240/Product: superoxide dismutase (Cu-Zn), extracellular #status predicted

F:107/Binding site: carbohydrate (Asn) (covalent) #status predicted

F:114,116,131,181/Binding site: copper (His) #status predicted

F:125-208/disulfide bonds: #status predicted

F:131,139,142,145/Binding site: zinc (His, His, His, His, Asp) #status predicted

F:204/Active site: Arg #status predicted

Query Match 1.8%; Score 7; DB 1; Length 240;

Best Local Similarity 100.0%; Pred. No. 60;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 358 PSATLDA 364

Db 67 PSATLDA 73

RESULT 21

A83799

hypothetical protein BH1193 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 15-Jun-2001

C:Accession: A83799

R: Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: A83799

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-249 <STO>

A:Cross-references: GB:AP001511; GB:BA000004; NID:g10173727; PIDN:BA04912.1; GSPDB:GN00

C:Genetics:

A:Gene: BH1193

Query Match 1.8%; Score 7; DB 2; Length 249;

Best Local Similarity 100.0%; Pred. No. 63;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 65 YEEQIQE 71

Db 4 YEEQIQE 10

RESULT 22

B90421

hypothetical protein SS02489 [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 24-May-2001

C:Accession: B90421

R: She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-I

Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.

arratt, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.

submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: B90421

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-254 <KUR>

A:Cross-references: GB:AE006641; NID:g13815792; PIDN:AAK42625.1; GSPDB:GN00155

C:Genetics:

A:Gene: SS02489

Query Match 1.8%; Score 7; DB 2; Length 254;

Best Local Similarity 100.0%; Pred. No. 64;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 86 LIPSLGV 92

Db 176 LIPSLGV 182

RESULT 23

T33405

hypothetical protein H10E21.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 04-Mar-2000

C:Accession: T33405

R: Davidson, S.; Wohldmann, P.; Courtney, L.

submitted to the EMBL Data Library, July 1998

A:Description: The sequence of C. elegans cosmid H10E21.

A:Reference number: Z21339

A:Accession: T33405

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-260 <DAV>

A:Cross-references: EMBL:AF078783; PIDN:AAC36919.1; GSPDB:GN000021; CESP:H10E21.1

A:Experimental source: strain Bristol N2; clone H10E21

C:Genetics:

A:Gene: CESP:H10E21.1

A:Map position: 3

A:Introns: 48/3; 119/1; 160/1; 217/3

C:Superfamily: Caenorhabditis elegans hypothetical protein H10E21.1

Query Match 1.8%; Score 7; DB 2; Length 260;

Best Local Similarity 100.0%; Pred. No. 65;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 124 ISROINS 130  
|||||  
Db 190 ISROINS 196

## RESULT 24

A82304  
3-methyl-2-oxobutanoate hydroxymethyltransferase VC0592 [imported] - Vibrio cholerae (sp  
C/Species: Vibrio cholerae  
C/Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C/Accession: A82304  
R/Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gilm, M.L.; Dodson, R.U.;  
Chardon, D.; Ermolova, M.D.; Vamathevan, J.; Baas, S.; Qin, H.; Dragoi, I.; Sellers, E.  
1, R.R.; Mekalanos, J.D.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A/Reference number: A82035, NCID:20406833, PMID:10952301  
A/Accession: A82304  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-264 <HEI>  
A/Cross-references: GB:AE004144; GB:AE003852; NID:99655017; PIDN:AAF93759.1; GSPDB:GN001  
A/Experimental source: serogroup O1, strain N16961, biotype El Tor  
A/Genetic: VC0592  
A/Map position: 1  
C:Superfamily: 3-methyl-2-oxobutanoate hydroxymethyltransferase

Query Match 1.8%; Score 7; DB 2; Length 264;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 50 LANRITQ 56  
|||||  
Db 187 LANRITQ 193

## RESULT 25

B98179  
probable NAD(P)H nitroreductase yfko [imported] - Agrobacterium tumefaciens (strain CS8,  
C/Species: Agrobacterium tumefaciens  
C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
C/Accession: B98179  
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Woliam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Marzelz, B.;  
Science 294, 2323-2328, 2001  
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
A/Reference number: A97359; PMID:11743194  
A/Accession: B98179  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-271 <KUR>  
A/Cross-references: GB:AE007870; PIDN:AAK88956.1; PID:gl5158735; GSPDB:GN00170  
C/Genetic:  
A/Map position: linear chromosome

Query Match 1.8%; Score 7; DB 2; Length 271;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 SUPPLSV 17  
|||||  
Db 9 SUPPLSV 15

## RESULT 26

C69030  
MJ1225 protein homolog MTH122 - Methanobacterium thermoautotrophicum (strain Delta H)  
N:Alternate names: inosine-5'-monophosphate dehydrogenase related protein I [minimer]  
C/Species: Methanobacterium thermoautotrophicum  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000  
C/Accession: C69030

R/Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;  
Qiu, D.; Spadofora, R.; Vicalire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N  
ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A/Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: func  
A/Reference number: A69000; NCID:98037514; PMID:9377463  
A/Accession: C69030  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-275 <MTH>

A/Cross-references: GB:AE000890; GB:AE000666; NID:92622331; PIDN:AA85711.1; PID:9262233  
A/Experimental source: strain Delta H  
C/Genetic:  
A/Genetic: MTH1222  
A/Start codon: GTG  
C:Superfamily: conserved hypothetical protein MJ1225; CBS homology  
C/Keyword: duplication  
F:13-65/Domain: CBS homology <CBS1>  
F:86-133/Domain: CBS homology <CBS2>  
F:149-196/Domain: CBS homology <CBS3>  
F:223-269/Domain: CBS homology <CBS4>

Query Match 1.8%; Score 7; DB 1; Length 275;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 160 IYTKSDL 166  
|||||  
Db 122 IYTKSDL 128

## RESULT 27

G86269  
protein P21P23.9 [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C/Accession: G86269  
R/Theologis, A.; Ecker, J.R.; Palm, C.D.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
angen, N.F.; Hughes, B.; Huizart, L.  
Nature 408, 816-820, 2000  
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Talion,  
ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Reference number: A86141; NCID:21016719; PMID:11130712  
A/Accession: G86269  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-281 <STO>  
A/Cross-references: GB:AE005172; NID:98920570; PIDN:AAF81292.1; GSPDB:GN00141  
C/Genetic:  
A/Map position: 1

Query Match 1.8%; Score 7; DB 2; Length 281;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 348 ALPILQS 354  
|||||  
Db 160 ALPILQS 166

## RESULT 28

B85709  
unknown protein encoded by prophage CP-9330 [imported] - Escherichia coli (strain O157:H  
C/Species: Escherichia coli  
C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C/Accession: B85709  
R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaeser, J.D.; Rose, D.J.; Mayhew

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca, Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
A:Reference number: AB5480; MUID:21074935; PMID:11206551  
A:Accession: B95709  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-290 <STO>  
A:Cross-references: GB:AB005174; NID:gl21515010; PIDN:AAG56137.1; GSPDB:GN00145; UWGP:220  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: Z2066

Query Match 1.8%; Score 7; DB 2; Length 290;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 275 DYGNRLA 281  
|||||  
DB 232 DYGNRLA 238

RESULT 29  
A99818  
hypothetical protein SA0470 [imported] - *Staphylococcus aureus* (strain N315)  
C:Species: *Staphylococcus aureus*  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: A99818  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiranatsu, K.  
Lancet 357, 1235-1240, 2001  
A:Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.  
A:Reference number: A99758; MUID:21311952; PMID:11418146  
A:Accession: A99818  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-293 <KUR>  
A:Cross-references: GB:BA000018; PID:gl3700402; PIDN:BA041700.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA0470  
C:Superfamily: conserved hypothetical protein sl1198

Query Match 1.8%; Score 7; DB 2; Length 293;  
Best Local Similarity 100.0%; Pred. No. 73;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 290 LNAIKGL 296  
|||||  
DB 246 LNAIKGL 252

RESULT 30  
A13635  
thiamine biosynthesis lipoprotein appE precursor [imported] - *Brucella melitensis* (strain  
C:Species: *Brucella melitensis*  
C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 15-Feb-2002  
C:Accession: A13635  
R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,  
.; Mazur, M.; Goldsman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letess  
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*  
A:Reference number: AB3252; PMID:11756688  
A:Accession: A13635  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-325 <KUR>  
A:Cross-references: GB:AE008918; PIDN:AAL54252.1; PID:gl17985225; GSPDB:GN00191  
A:Experimental source: strain 16M  
C:Genetics:  
A:Gene: BME11010  
A:Map position: 11

C:Superfamily: hypothetical protein HI0172

Query Match 1.8%; Score 7; DB 2; Length 325;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 TTGGAFDA 221  
|||||  
DB 93 TTGGAFDA 99

# RESULT 31

H95342

nosX protein required for nitrous oxide reduction [imported] - *Sinorhizobium meliloti* (s  
C:Species: *Sinorhizobium meliloti*  
C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
C:Accession: H95342  
R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bowe  
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C.  
Proc. Natl. Acad. Sci. U.S.A. 98, 9893-9898, 2001  
A:Title: Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti*  
A:Reference number: A95262; MUID:21396509; PMID:11481432  
A:Accession: H95342  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-334 <KUR>  
A:Cross-references: GB:AE006469; PIDN:AAK65306.1; PID:gl14523761; GSPDB:GN00165  
A:Experimental source: strain 1021, megaplasmid pSymA  
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001  
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
hebaull, P.; Vandenberg, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.  
A:Reference number: A96039; MUID:21368234; PMID:11474104  
A:Contents: annotation  
C:Genetics:  
A:Gene: nosX  
A:Gene: plasmid

Query Match 1.8%; Score 7; DB 2; Length 334;  
Best Local Similarity 100.0%; Pred. No. 83;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 TTGGAFD 220  
|||||  
DB 115 TTGGAFD 121

# RESULT 32

B97688

C4-dicarboxylate transport system (C4-dicarboxylate-binding protein) (AP001516) [imported  
C:Species: *Agrobacterium tumefaciens*  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
C:Accession: B97688  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A:Title: Genome sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens*  
A:Reference number: A97359; PMID:11743194  
A:Accession: B97688  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-335 <KUR>  
A:Cross-references: GB:AE007869; PIDN:AAK88459.1; PID:gl15157562; GSPDB:GN00169  
C:Genetics:  
A:Gene: AGP C\_4976

Query Match 1.8%; Score 7; DB 2; Length 335;  
Best Local Similarity 100.0%; Pred. No. 83;

```

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 86 LIPSLGV 92
    |||||
Db 99 LIPSLGV 105

RESULT 33
AG2913
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AG2913
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo, I.
erage, G.; Gillet, W.; Grant, C.; Genthner, D.; Kuyavian, T.; Levy, R.; Li, M.; McClell
i, Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
Ber, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AG2913
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-335 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL43725.1; PID:g17741256; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
A:Gene: Atu2744
A:Map position: circular chromosome
C:Superfamily: conserved hypothetical protein H1028

Query Match 1.8%; Score 7; DB 2; Length 335;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 86 LIPSLGV 92
    |||||
Db 99 LIPSLGV 105

RESULT 34
C82190
format dehydrogenase accessory protein VC1519 [imported] - Vibrio cholerae (strain N169
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: C82190
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoli, I.; Sellers, F.
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; MUID:20406833; PMID:10952301
A:Accession: C82190
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337 <HEI>
A:Cross-references: GB:AE004230; GB:AE003852; NID:g9656018; PIDN:AAF94673.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; Diocyte EI For
A:Gene: VC1519
A:Map position: 1

Query Match 1.8%; Score 7; DB 2; Length 337;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 260 ALTHYND 266
    |||||
Db 208 ALTHYND 214

RESULT 35

```

```

E75436
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: E75436
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Ulteback, T.; Zalewski, C.; M
S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Ulteback, T.; Zalewski, C.; M
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567286
A:Accession: E75436
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-340 <WHI>
A:Cross-references: GB:AE001961; GB:AE00513; NID:g6456843; PIDN:AAF10688.1; PID:g645685
A:Experimental source: strain R1
A:Gene: DR118
A:Map position: 1
C:Superfamily: conserved hypothetical protein yacL

Query Match 1.8%; Score 7; DB 2; Length 340;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 14 LLSVAVT 20
    |||||
Db 107 LLSVAVT 113

RESULT 36
AH3338
cobw protein [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 17-May-2002
C:Accession: AH3338
R:DelVecchio, V.G.; Kapral, V.; Redkar, R.D.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AH3338
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL51875.1; PID:g17982626; GSPDB:GN00190
A:Experimental source: strain 16M
A:Gene: BMRI0694
A:Map position: I
C:Superfamily: cobw protein

Query Match 1.8%; Score 7; DB 2; Length 349;
Best Local Similarity 100.0%; Pred. No. 86;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 221 ANGKRIA 227
    |||||
Db 30 ANGKRIA 36

RESULT 37
AD0374
Phosphoribosylaminimidazole carboxylase (EC 4.1.1.21) ATPase chain Purk [imported] - Ye
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 11-Jan-2002
C:Accession: AD0374
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tlhal, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

```

A;Reference number: AB0001; MUID:21470413; PMID:11586360  
A;Accession: AD0374  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-354 <KUR>  
A;Cross-references: GB:AL590842; PIDN:CAC92319.1; PID:gl5981030; GSPDB:GN00175  
C;Genetics:  
A;Gene: purK  
C;Superfamily: phosphoribosylaminoimidazole carboxylase carbon dioxide-fixation chain; P  
C;Keywords: carbon-carbon lyase; carboxy-lyase

Query Match 1.8%; Score 7; DB 2; Length 354;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 8 QWLSLPL 14  
Db 293 QWLSLPL 299

RESULT 38  
B70189  
rod shape-determining protein (mreB-1) homolog - Lyme disease spirochete  
C;Species: Borrelia burgdorferi (Lyme disease spirochete)  
C;Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 26-Aug-1999  
C;Accession: B70189  
R;Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White  
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,  
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A;Authors: Smith, H.O.; Venter, J.C.  
A;Title: Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.  
A;Reference number: A70100; MUID:98065943; PMID:9403685  
A;Accession: B70189  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-361 <KLE>  
A;Cross-references: GB:AE001171; GB:AE000783; NID:g2688640; PIDN:AAAC67059.1; PID:g268864  
A;Experimental source: strain B31  
C;Superfamily: rod shape-determining protein envB

Query Match 1.8%; Score 7; DB 2; Length 361;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 13 PLLSVAV 19  
Db 331 PLLSVAV 337

RESULT 39  
E96904  
m1ND family ATPase from Para/SOJ subfamily [imported] - Clostridium acetobutylicum  
C;Species: Clostridium acetobutylicum  
C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 14-Sep-2001  
C;Accession: E96904  
R;Nolling, J.; Breton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
J.; Dally, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl  
A;Reference number: A96900; MUID:21359325; PMID:21359325  
A;Accession: E96904  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-361 <KUR>  
A;Cross-references: GB:AE001437; PIDN:AAK78024.1; PID:gl5022858; GSPDB:GN00168  
A;Experimental source: Clostridium acetobutylicum ATCC824  
C;Genetics:  
A;Gene: CAC0037

Query Match 1.8%; Score 7; DB 2; Length 361;  
Best Local Similarity 100.0%; Pred. No. 89;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 307 STFNINY 313  
Db 254 STFNINY 260

## RESULT 40

AF1970

3-isopropylmalate dehydrogenase [imported] - Nostoc sp. (strain PCC 7120)  
C;Species: Nostoc sp.

A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C;Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002

C;Accession: AF1970  
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,  
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.  
DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AF1970

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-362 &lt;KUR&gt;

A;Cross-references: GB:BA000019; PIDN:BA873270.1; PID:gl7130660; GSPDB:GN00179

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: leuB

C;Superfamily: 3-isopropylmalate dehydrogenase

Query Match 1.8%; Score 7; DB 2; Length 362;

Best Local Similarity 100.0%; Pred. No. 89;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 357 LPSATLD 363

Db 57 LPSATLD 63

## RESULT 41

A41601

Na+/taurocholate transport protein - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 16-Feb-1997

C;Accession: A41601

R;Hagenbuch, B.; Stieger, B.; Foguet, M.; Luebbert, H.; Meier, P.J.

Proc. Natl. Acad. Sci. U.S.A. 88, 10629-10633, 1991

A;Title: Functional expression cloning and characterization of the hepatocyte Na(+)/bile  
A;Reference number: A41601; MUID:92073340; PMID:1961729

A;Accession: A41601

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-362 &lt;HAG&gt;

A;Cross-references: GB:M77429

C;Keywords: transmembrane protein

Query Match 1.8%; Score 7; DB 2; Length 362;

Best Local Similarity 100.0%; Pred. No. 89;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 LLSVAVT 20

Db 197 LLSVAVT 203

## RESULT 42

C90646

probable fimbrial protein [imported] - Escherichia coli (strain O157:H7, substrain RMD  
C;Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001

C;Accession: C90646

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene

A:Reference number: A99629; MUID:21156231; PMID:11258796  
 A:Accession: C90646  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-369 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BAB33562.1; PID:q13359595; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain R1MD 0509952  
 C:Genetics:  
 A:Gene: ECG0139

Query Match 1.8%; Score 7; DB 2; Length 369;  
 Best Local Similarity 100.0%; Pred. No. 91;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GVSSGTT 97  
 |||||  
 Db 184 GVSSGTT 190

RESULT 43  
 C85497  
 probable fimbrial protein yadC [imported] - Escherichia coli (strain O157:H7, substrain  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
 C:Accession: C85497  
 R:Perera, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.U.; Mayhew  
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoustis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: A85480; MUID:21074935; PMID:11206551  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: yadC  
 A:Molecule type: DNA  
 A:Status: preliminary  
 A:Accession: C85497  
 A:Residues: 1-369 <STO>  
 A:Cross-references: GB:AB005174; NID:g12512853; PIDN:AAG54439.1; GSPDB:GN00145; UWGP:201  
 A:Experimental source: strain O157:H7, substrain EDL933  
 C:Genetics:  
 A:Gene: yadC

Query Match 1.8%; Score 7; DB 2; Length 369;  
 Best Local Similarity 100.0%; Pred. No. 91;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 91 GVSSGTT 97  
 |||||  
 Db 184 GVSSGTT 190

RESULT 44  
 GNVSMB  
 genome polypeptide - maize dwarf mosaic virus (strain B) (fragment)  
 N:Contains: carboxyl end of nuclear inclusion protein b; coat protein  
 C:Species: maize dwarf mosaic virus, MDWV  
 C:Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 17-Nov-2000  
 C:Accession: PH0208  
 R:Frenkel, M.J.; Jilka, J.M.; McKern, N.M.; Strike, P.M.; Clark Jr., J.M.; Shukla, D.D.;  
 J. Gen. Virol. 72, 237-242, 1991  
 A:Title: Unexpected sequence diversity in the amino-terminal ends of the coat proteins c  
 A:Reference number: PH0207; MUID:91132116; PMID:1993866  
 A:Accession: PH0208  
 A:Molecule type: genomic RNA  
 A:Residues: 1-380 <RRS>  
 A:Cross-references: GB:D00949; NID:g222059; PIDN:BA00797.1; PID:g222060  
 C:Superfamily: tobacco etch virus genome polypeptide  
 C:Keywords: coat protein; inclusion protein b (fragment) #status predicted <NIP>  
 F:1-52/Product: nuclear inclusion protein b (fragment) #status predicted <COA>  
 F:53-380/Product: coat protein #status predicted <COA>

Query Match 1.8%; Score 7; DB 1; Length 380;  
 Best Local Similarity 100.0%; Pred. No. 93;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 GSGATG 156

Db 111 GSGATG 117  
 |||||

RESULT 45  
 B90682  
 probable enzyme [imported] - Escherichia coli (strain O157:H7, substrain R1MD 0509952)  
 C:Species: Escherichia coli  
 C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 24-Aug-2001  
 C:Accession: B90682  
 R:Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
 Gaeawa, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.  
 DNA Res. 8, 11-22, 2001  
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc  
 A:Reference number: A99629; MUID:21156231; PMID:11258796  
 A:Accession: B90682  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-385 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BA33849.1; PID:q13359883; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain R1MD 0509952  
 C:Genetics:  
 A:Gene: ECG0426  
 C:Superfamily: Escherichia coli beta-lactamase

Query Match 1.8%; Score 7; DB 2; Length 385;  
 Best Local Similarity 100.0%; Pred. No. 94;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 YGSGATG 155  
 |||||  
 Db 44 YGSGATG 50

Search completed: December 25, 2002, 20:32:44  
 Job time : 41 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 25, 2002, 20:22:58 / Search time 17 Seconds  
(without alignments)  
975.914 Million cell updates/sec

Title: US-09-889-746-2  
Perfect score: 400  
Sequence: 1 MHSHPQWLSLPLSLVAVT.....SKAEVLGRVFNKPKRALF 400

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10	2.5	732	1	IUTR_ECOLI
2	8	2.0	465	1	FXD3_MOUSE
3	8	2.0	895	1	DAG1_BOVIN
4	7	1.8	62	1	NXS2_LATCO
5	7	1.8	62	1	NXSC_LATCO
6	7	1.8	62	1	NXSD_LATCO
7	7	1.8	198	1	GRPE_CAUCR
8	7	1.8	240	1	SODE_HUMAN
9	7	1.8	244	1	SODE_RABIT
10	7	1.8	252	1	1431_SCMA
11	7	1.8	264	1	PANB_VIBCH
12	7	1.8	362	1	NTCP_RAT
13	7	1.8	367	1	GP17_HUMAN
14	7	1.8	380	1	POLG_MDMV
15	7	1.8	385	1	AMPH_ECOLI
16	7	1.8	426	1	MUTL_AOUPY
17	7	1.8	433	1	OADB_SALT1
18	7	1.8	433	1	OADB_SALT1
19	7	1.8	457	1	LEF4_NPOVP
20	7	1.8	462	1	VS11_REBOV
21	7	1.8	474	1	UCR2_EUGCR
22	7	1.8	576	1	MUTL_CHLTR
23	7	1.8	617	1	IF2_MYCPN
24	7	1.8	648	1	PTAA_ECOLI
25	7	1.8	746	1	RHTA_RHIME
26	7	1.8	793	1	CLPA_RHOBL
27	7	1.8	806	1	NPVAC
28	7	1.8	1157	1	PEX1_PICPA
29	6	1.5	20	1	RIPY_CUCPE
30	6	1.5	33	1	PEN3_ADECU
31	6	1.5	45	1	RIP2_TRIKI
32	6	1.5	46	1	ILVD_RHOCA
33	6	1.5	55	1	YME2_THIFE

34	6	1.5	56	1	VG1B_BPPZA	P06948 bacterioph
35	6	1.5	66	1	YSCB_YEREN	Q01246 yersinia en
36	6	1.5	79	1	NSGX_HUMAN	O9uh64 homo sapien
37	6	1.5	83	1	UB2_PLAPE	P21857 platichtyvs
38	6	1.5	86	1	Y16_BPT7	P03793 bacterioph
39	6	1.5	94	1	ITFB_BUCAI	P57394 buchnera ap
40	6	1.5	94	1	YDFP_SCHPO	Q10486 schizosacch
41	6	1.5	96	1	Y1LK_BPCHP	P19183 bacterioph
42	6	1.5	98	1	YACB_RHISN	P55384 rhizobium s
43	6	1.5	105	1	YJNS_YEAST	P47012 saccharomyc
44	6	1.5	111	1	SS18_STRYG	P80388 streptomyc
45	6	1.5	111	1	YCK5_OENHO	Q9mct3 oenotiera h
46	6	1.5	118	1	Y13X_HCVVA	P03236 human cytom
47	6	1.5	124	1	UR2_HUMAN	O95399 homo sapien
48	6	1.5	125	1	YRB2_CAEEL	O09399 caenorhabdi
49	6	1.5	128	1	AZUR_PSECT	P00286 pseudomonas
50	6	1.5	131	1	PRDC_PROTU	P20923 proteus vul
51	6	1.5	134	1	RR9_EUGGR	P32060 euglena gra
52	6	1.5	135	1	Y40A_TREPA	P56007 treponema p
53	6	1.5	137	1	RL16_SPICT	O31162 spiroplasma
54	6	1.5	138	1	RR9_ODOSI	P49497 odontella s
55	6	1.5	142	1	RS16_TORBU	O9xek7 tortulia rur
56	6	1.5	145	1	GTCL_LISMO	P56412 listeria mo
57	6	1.5	145	1	GTCA_LISIN	O92704 listeria in
58	6	1.5	145	1	PLAS_CHLRE	P18068 chlamydomon
59	6	1.5	145	1	RS16_FRIG	O22647 fritillaria
60	6	1.5	145	1	RS16_GOSHI	P46293 gossypium h
61	6	1.5	145	1	RS16_LUPRO	P16149 lupinus pol
62	6	1.5	145	1	VG10_BPMIS	Q05284 mycobacteri
63	6	1.5	146	1	RS16_ARAVH	Q42340 atrebiopsis
64	6	1.5	151	1	RK28_TOBAC	P30956 nicotiana t
65	6	1.5	153	1	ELAA_ECOLI	P52077 escherichia
66	6	1.5	153	1	THYV_RANCA	P31779 rana catesb
67	6	1.5	154	1	AAU3_RHIME	O9x9m8 rhizobium m
68	6	1.5	154	1	YCS6_HAEIN	P44142 haemophilus
69	6	1.5	157	1	KNMB_MAIZE	P27347 zea mays (m
70	6	1.5	160	1	EZRI_RAT	P31977 rattus norv
71	6	1.5	161	1	THYL_CHICK	Q07212 gallus gall
72	6	1.5	161	1	HMGL_WHEAT	P40621 triticum ae
73	6	1.5	162	1	Y415_METJA	Q57858 methanococc
74	6	1.5	169	1	FLAV_SYNP2	P31158 synecococc
75	6	1.5	170	1	YGR5_YEAST	P46591 saccharomyc
76	6	1.5	171	1	FABA_RHIL0	O9ebh8 rhizobium l
77	6	1.5	172	1	FABA_BRUME	O9ebc2 bruceella me
78	6	1.5	172	1	GP38_MOUSE	Q62011 mus musculu
79	6	1.5	176	1	TRAF_AGRIS	Q44350 agrobacteri
80	6	1.5	178	1	YCS6_HAEIN	Q57519 haemophilus
81	6	1.5	179	1	Y281_METJA	Q57729 methanococc
82	6	1.5	182	1	YLP2_CAEEL	P34382 caenorhabdi
83	6	1.5	183	1	CITX_ECOLI	P7563 escherichia
84	6	1.5	183	1	Y829_ARCRU	O29429 archaeoglob
85	6	1.5	185	1	YCOC_ECOLI	P38522 escherichia
86	6	1.5	186	1	RS7_METTH	O27130 methanobact
87	6	1.5	188	1	KTHY_METJA	O57741 methanococc
88	6	1.5	190	1	CASK_CERNI	P42157 cervinus nldp
89	6	1.5	190	1	SARI_SCHPO	Q01475 schizosacch
90	6	1.5	197	1	YUT5_ANASP	P56632 arabidopsi
91	6	1.5	198	1	RR27_ARAYH	O9f1n4 arabidopsi
92	6	1.5	198	1	YV47_YEAST	Q03233 saccharomyc
93	6	1.5	199	1	MI03_SCHPO	Q09330 schizosacch
94	6	1.5	199	1	PEXH_YEAST	P40155 saccharomyc
95	6	1.5	199	1	TDX2_BRUMA	Q17172 brugia mala
96	6	1.5	202	1	LSPA_MYCTU	O10764 mycobacteri
97	6	1.5	203	1	PAK1_CHICK	P47236 gallus gall
98	6	1.5	203	1	SOMA_PAGUA	P08591 pagrus mayo
99	6	1.5	204	1	SOMA_AACABU	Q01282 acanthopagr
100	6	1.5	204	1	SOMA_AACALA	P45654 acanthopagr
101	6	1.5	204	1	SOMA_LATCA	Q01283 acanthopagr
102	6	1.5	204	1	SOMA_MORSA	P48428 morone saxa
103	6	1.5	204	1	SOMA_PERRY	O9ev43 petra flavo
104	6	1.5	204	1	SOMA_SCIOC	O9ib11 sciaenops o
105	6	1.5	204	1	SOMA_SEBSC	P87391 sebastes ac
106	6	1.5	204	1	SOMA_SEROU	P08539 seriola qui

107	6	1.5	204	1	SOMA_SPRAU	P29971	sparus aura	180	6	1.5	290	1	SGCG_HUMAN	Q13326	homo sapien
108	6	1.5	204	1	SOMA_THUTH	P09113	thunus thy	181	6	1.5	292	1	COX2_AQUAE	O66544	aquifex ao
109	6	1.5	204	1	SOMA_TRITC	P29uf6	trichogaste	182	6	1.5	293	1	LEC1_CLALU	Q95284	cladrasia
110	6	1.5	205	1	Y511_HALMA	P20191	haloarcula	183	6	1.5	294	1	PPNK_VIBCH	Q39t08	vibriol chol
111	6	1.5	206	1	KTHY_DEIRA	Q9ty40	deinococcus	184	6	1.5	295	1	AQP9_RAT	P56677	rattus norv
112	6	1.5	206	1	RALB_HUMAN	P11234	homo sapien	185	6	1.5	295	1	MPR1_SCHPO	O94321	achisovacch
113	6	1.5	206	1	RALB_RAT	P36860	rattus norv	186	6	1.5	296	1	ZA2G_RAT	Q83678	rattus norv
114	6	1.5	206	1	YGIM_ECOLI	P39202	escherichia	187	6	1.5	297	1	OE27_NPOVP	O10372	orgyia paeu
115	6	1.5	207	1	GIDB_SALTU	Q9xfw0	salmonella	188	6	1.5	299	1	HIS1_ECO57	Q8x8t4	escherichia
116	6	1.5	209	1	IM23_HUMAN	O14925	homo sapien	189	6	1.5	299	1	HIS1_ECOLI	P10366	escherichia
117	6	1.5	209	1	IM23_MOUSE	Q9wtq8	mus musculus	190	6	1.5	300	1	SDB1_RAT	Q9ji92	rattus norv
118	6	1.5	209	1	IM23_RAT	Q35093	rattus norv	191	6	1.5	301	1	YGRJ_HAEIN	P44979	haemophilus
119	6	1.5	211	1	WFD1_MOUSE	Q9esh5	mus musculus	192	6	1.5	302	1	DEVK_MYXXA	P49765	myxococcus
120	6	1.5	212	1	WFD1_RAT	Q70280	rattus norv	193	6	1.5	304	1	HIS1_VIBCH	Q9kxk4	vibriol chol
121	6	1.5	213	1	CPCF_SYNEL	P50038	synechococc	194	6	1.5	307	1	MYIM_ECOLI	P75931	escherichia
122	6	1.5	215	1	KAD_NEIGO	P49979	neisseria g	195	6	1.5	307	1	PABL_STRGR	P33585	streptomyce
123	6	1.5	215	1	KAD_NEIMA	P49980	neisseria m	196	6	1.5	307	1	YI23_MYCTU	Q50610	mycobacteri
124	6	1.5	215	1	TAL_CLOAB	Q97jd9	clostridium	197	6	1.5	307	1	ZA2G_MOUSE	Q64726	mus musculus
125	6	1.5	215	1	YG35_YEAST	P53277	saccharomyc	198	6	1.5	312	1	MTM4_NEIGO	P31033	neisseria g
126	6	1.5	216	1	A674_CHVFI	O41156	paramecium	199	6	1.5	312	1	VNEM_EBV	P03208	epstein-bar
127	6	1.5	218	1	CIP2_STRCO	Q9zh58	streptomyce	200	6	1.5	315	1	FLIM_RHIME	O54246	thizobium m
128	6	1.5	222	1	IM23_YEAST	P32897	saccharomyc	201	6	1.5	316	1	YK95_MYCTU	Q10704	mycobacteri
129	6	1.5	222	1	YW82_MYCTU	P96887	mycobacteri	202	6	1.5	317	1	YK05_YEAST	P25616	saccharomyc
130	6	1.5	225	1	RNC_VIBCH	Q9kpb2	vibriol chol	203	6	1.5	318	1	YIL7_YEAST	P40476	saccharomyc
131	6	1.5	227	1	COX2_CHOFI	P98025	christoneu	204	6	1.5	319	1	CATL_SCHMA	Q26534	echistocoma
132	6	1.5	227	1	COX2_CHOFU	P98026	christoneu	205	6	1.5	319	1	METX_METH	Q7848	methanobact
133	6	1.5	227	1	COX2_CHORO	P98030	christoneu	206	6	1.5	321	1	FMT_CHLPN	P20746	avian adeno
134	6	1.5	227	1	COX2_EULMA	P98033	eulem mac	207	6	1.5	321	1	YOR4_ADEG1	Q927Q5	chlamydia p
135	6	1.5	228	1	COX2_DINSE	O79549	dinodon sem	208	6	1.5	323	1	ANX5_CYNPY	P70075	cynopy pyrr
136	6	1.5	228	1	COX2_GALME	P29874	galleria me	209	6	1.5	324	1	YK95_MYCLE	P54075	mycobacteri
137	6	1.5	228	1	COX2_PBRAM	P29877	periplaneta	210	6	1.5	324	1	YMO2_MYCTU	Q10331	mycobacteri
138	6	1.5	229	1	COX2_ONCPA	P29876	oncopeltus	211	6	1.5	328	1	APPD_BACSU	P42064	bacillus su
139	6	1.5	232	1	CASB_PIG	P39037	sus scrofa	212	6	1.5	329	1	QOR_HUMAN	Q08257	homo sapien
140	6	1.5	234	1	SYM_BUCAP	Q9zhd7	buchnera ap	213	6	1.5	329	1	YI46_HAEIN	P44542	haemophilus
141	6	1.5	235	1	PUR7_CLOPE	O8xmk6	clostridium	214	6	1.5	330	1	VP19_HSVSA	Q10511	herpesvirus
142	6	1.5	237	1	Y511_TREPA	O83732	treponema p	215	6	1.5	337	1	D3H1_ARATH	Q9auc0	arabidopsis
143	6	1.5	239	1	Y559_YEAST	P53314	saccharomyc	216	6	1.5	337	1	FABH_PSEAE	P20582	pseudomonas
144	6	1.5	242	1	Y755_AQUAE	O66957	aquifex ao	217	6	1.5	337	1	G3PC_ANTMA	P25861	antirrhinum
145	6	1.5	245	1	CTRA_BOVIN	P00766	bos taurus	218	6	1.5	337	1	G3PC_PETHY	P26520	petunia hyb
146	6	1.5	248	1	PFLA_LISMO	Q9x767	listeria mo	219	6	1.5	337	1	G3PC_SINAL	P04796	sinapis alb
147	6	1.5	248	1	TPM4_RAT	P09495	rattus norv	220	6	1.5	337	1	G3PD_MAIZE	Q09054	zea mays (m
148	6	1.5	248	1	Y964_PSEAE	Q51423	pseudomonas	221	6	1.5	337	1	G3PX_HORVU	P26517	hordeum vul
149	6	1.5	250	1	PSM4_CAEL	Q9n599	caenorhabdi	222	6	1.5	337	1	YAJ2_PSEAE	Q9hve4	pseudomonas
150	6	1.5	253	1	TSFY_HUMAN	Q01534	homo sapien	223	6	1.5	338	1	BZTM_RHOCA	Q52663	rhodobacter
151	6	1.5	256	1	YGBI_HAEIN	P44978	haemophilus	224	6	1.5	338	1	CYP_PHOLA	P95522	phormidium
152	6	1.5	259	1	COX3_LOLBI	O47475	loligo ble	225	6	1.5	338	1	G3PC_RANAC	P25858	arabidopsis
153	6	1.5	259	1	COX3_LUMTE	Q34943	lumbrius t	226	6	1.5	338	1	RLA0_SULSO	P26521	tanunculus
154	6	1.5	260	1	PSA1_CAEL	O44156	caenorhabdi	227	6	1.5	338	1	PAX9_CHICK	P56039	gallus gall
155	6	1.5	260	1	VG38_BPK3	P07876	bacterioph	228	6	1.5	339	1	PAX9_HUMAN	P55166	gallus gall
156	6	1.5	261	1	6PGL_STRCO	Q9xab7	streptomyce	229	6	1.5	339	1	Y923_TREPA	O83893	treponema p
157	6	1.5	261	1	YHC4_CAEL	Q23314	caenorhabdi	230	6	1.5	340	1	ARGC_STRCL	P54896	streptomyce
158	6	1.5	262	1	VG38_BPT2	P07875	bacterioph	231	6	1.5	340	1	TDH_XANCP	O34268	xanthomonas
159	6	1.5	263	1	CLPE_ECOLI	Q05433	escherichia	232	6	1.5	341	1	G3PC_MAGLI	P26518	magnolia li
160	6	1.5	263	1	ONCM_MOUSE	P53347	mus musculus	233	6	1.5	341	1	OXDD_BOVIN	P31228	bos taurus
161	6	1.5	266	1	OTOF_RAT	Q9erc5	rattus norv	234	6	1.5	341	1	OXDD_HUMAN	Q99489	homo sapien
162	6	1.5	268	1	YJ80_AERPE	Q9yaf9	aeropyrum p	235	6	1.5	341	1	PAX9_HUMAN	P55771	homo sapien
163	6	1.5	269	1	RF4Z_SALTU	P26473	salmonella	236	6	1.5	342	1	PAX9_MOUSE	P47242	mus musculus
164	6	1.5	269	1	SRPB_MOUSE	P47758	mus musculus	237	6	1.5	342	1	SPDE_LYCES	Q92845	lycopersico
165	6	1.5	270	1	TRBG_RHISN	P55404	rhizobium s	238	6	1.5	343	1	ILVE_HAEIN	P54689	haemophilus
166	6	1.5	272	1	T2P1_HAEIN	P20588	haemophilus	239	6	1.5	346	1	DP42_MYCTU	P95102	mycobacteri
167	6	1.5	273	1	LECG_ARAHY	P02872	arachis hyp	240	6	1.5	347	1	JUND_HUMAN	P17535	homo sapien
168	6	1.5	274	1	STCQ_EMENT	Q00713	emeritella	241	6	1.5	349	1	DP04_PSEAE	Q91534	pseudomonas
169	6	1.5	275	1	BNZE_PSEPU	P08088	pseudomonas	242	6	1.5	350	1	UL88_HSV6U	P24438	human herpe
170	6	1.5	275	1	TODD_PSEPU	P13859	pseudomonas	243	6	1.5	350	1	YN04_YEAST	P53844	saccharomyc
171	6	1.5	276	1	FLJJ_CAUCR	P02969	caulobacter	244	6	1.5	352	1	YPBB_BACSU	P50728	bacillus su
172	6	1.5	280	1	BPBB_RHOGO	P47230	r cis-2,3-d	245	6	1.5	353	1	CHLI_GUITH	Q39516	guillardia
173	6	1.5	283	1	IPF1_RAT	P52947	rattus norv	246	6	1.5	353	1	CHLI_ODOSI	P49469	odontella s
174	6	1.5	283	1	RF4Z_ECOLI	P27241	escherichia	247	6	1.5	353	1	MURG_HELPJ	Q92K59	helicobacte
175	6	1.5	287	1	ID11_CLABR	Q39472	clarkia bre	248	6	1.5	353	1	MURG_HELPJ	Q52770	helicobacte
176	6	1.5	289	1	RIPS_TRIKI	P24478	trichosanth	249	6	1.5	354	1	CD68_HUMAN	P34810	homo sapien
177	6	1.5	289	1	RIPT_TRIKI	P09989	trichosanth	250	6	1.5	355	1	LAV1_PHYPO	P14725	physarum po
178	6	1.5	289	1	YF03_PASMU	Q9ckv2	pasteurella	251	6	1.5	357	1	AROC_PASMU	P57840	pasteurella
179	6	1.5	290	1	LPF4_AERHY	P45794	aeromonas h	252	6	1.5	357	1	MURG_PSEAE	Q9hw01	pseudomonas



253	6	1.5	357	1	TGN3_PAT	P19814	rattus norv	326	6	1.5	424	1	CRT2_ARATH	Q38858	arabidopsis
254	6	1.5	357	1	YCFT_ECOLI	P75955	escherichia	327	6	1.5	425	1	POXN_DROME	P23758	drosophila
255	6	1.5	358	1	YM05_MYCTU	Q10394	mycobacteri	328	6	1.5	425	1	TR16_RAT	P07174	rattus norv
256	6	1.5	359	1	MSAB_HELPJ	Q92kx8	h peptidase m	329	6	1.5	426	1	HISX_METTH	O26327	methanobact
257	6	1.5	359	1	MSAB_HELPJ	O25011	h peptidase m	330	6	1.5	427	1	TR16_HUMAN	P08138	homo sapien
258	6	1.5	359	1	THD1_SOLTU	P31612	solatium tub	331	6	1.5	429	1	HPAG_ECOLI	Q46978	e 4-hydroxy
259	6	1.5	360	1	CDXA_HUMAN	Q15131	homo sapien	332	6	1.5	429	1	HPAG_SALDU	Q92kx4	methanopyru
260	6	1.5	360	1	PURK_PSEAE	P72158	pseudomonas	333	6	1.5	431	1	SYD_METKA	O8cx94	methanopyru
261	6	1.5	360	1	RFL_AGRFS	Q8xbu8	agrobacteri	334	6	1.5	433	1	AFAR_EMENI	P53957	emericeella
262	6	1.5	360	1	RFL_PALSO	O8xxd0	raletonia s	335	6	1.5	433	1	IDHP_MEDSA	Q43435	medicago sa
263	6	1.5	361	1	LHX6_MOUSE	O88706	mus musculu	336	6	1.5	433	1	NCAP_PHV	P22047	prospect hi
264	6	1.5	361	1	PAX1_HUMAN	P15863	homo sapien	337	6	1.5	433	1	OADB_KLEPN	P13156	klebsiella
265	6	1.5	361	1	PAX1_MOUSE	P09084	mus musculu	338	6	1.5	433	1	OADB_VIBCH	Q9xtus	vibrio chol
266	6	1.5	362	1	NTCP_MOUSE	O08705	mus musculu	339	6	1.5	434	1	OADB_PASMU	Q96124	pasteurella
267	6	1.5	365	1	Y255_MYCGE	P47497	mycoplasma	340	6	1.5	434	1	PYRP_BACSU	P39765	bacillus su
268	6	1.5	366	1	OUQA_GAUCR	Q9at72	caulobacter	341	6	1.5	435	1	NPL6_YEAST	P32883	saccharomyc
269	6	1.5	370	1	ARIA_HUMAN	O92747	homo sapien	342	6	1.5	435	1	OADB_HAEPU	Q9rf98	haemophilus
270	6	1.5	370	1	HURK_RHILV	P28153	rhizobium l	343	6	1.5	437	1	VQ4_BPKE	P03667	bacteriopho
271	6	1.5	373	1	LYS1_YEAST	P38998	saccharomyc	344	6	1.5	442	1	RYK_AVIR3	P33497	avian retro
272	6	1.5	373	1	YOPO_BACSU	P54472	bacillus su	345	6	1.5	449	1	ARP3_YEAST	P47117	saccharomyc
273	6	1.5	374	1	TRM1_METUA	Q58356	methanococc	346	6	1.5	449	1	MANB_METJA	O57842	methanococc
274	6	1.5	375	1	FTSZ_HALN1	Q48290	halobacteri	347	6	1.5	449	1	YBB9_SCHPO	O60071	schizosacch
275	6	1.5	375	1	RDP_MOUSE	P19426	mus musculu	348	6	1.5	450	1	PAX8_HUMAN	Q06710	homo sapien
276	6	1.5	377	1	FLAF_VIBPA	O87081	vibrio para	349	6	1.5	451	1	GSHR_PSEAE	P23189	pseudomonas
277	6	1.5	380	1	FT22_METUA	Q58039	methanococc	350	6	1.5	451	1	PTK6_MOUSE	Q64434	mus musculu
278	6	1.5	380	1	RDP_HUMAN	P18615	homo sapien	351	6	1.5	455	1	DD15_STRPU	O17438	strongyloce
279	6	1.5	381	1	YFDE_ECOLI	P76518	escherichia	352	6	1.5	455	1	KR2_BEV	P13288	epstein-bar
280	6	1.5	386	1	METB_ECOLI	P00935	escherichia	353	6	1.5	456	1	DNAI_CHLMU	Q9pk64	chlamydia m
281	6	1.5	387	1	CTRA_NEIMA	P32758	neisseria m	354	6	1.5	456	1	DNAI_CHLTR	O84252	chlamydia t
282	6	1.5	389	1	ARGD_LACPL	O08321	lactobacill	355	6	1.5	456	1	GSHR_HAENI	P43783	haemophilus
283	6	1.5	391	1	CTRA_NEIMB	P32013	neisseria m	356	6	1.5	457	1	Y172_DROME	Q04283	drosophila
284	6	1.5	391	1	PAX5_HUMAN	Q02448	homo sapien	357	6	1.5	457	1	PAX8_MOUSE	O00288	mus musculu
285	6	1.5	391	1	PAX5_MOUSE	Q02650	mus musculu	358	6	1.5	457	1	PLSB_PEA	P30706	pium sativ
286	6	1.5	392	1	BCAM_HUMAN	O15382	homo sapien	359	6	1.5	457	1	Y4BP_RHISN	P55373	rhizobium s
287	6	1.5	392	1	FT22_ARCFU	Q29685	archaeoglob	360	6	1.5	458	1	PAX8_RAT	P51974	rattus norv
288	6	1.5	392	1	UDG5_ECOLI	Q47329	escherichia	361	6	1.5	459	1	N04M_CERAI	O03204	ceratotheri
289	6	1.5	393	1	YEC2_CAEEL	O45686	caenorhabdi	362	6	1.5	459	1	PAX8_CANPA	P47240	canis famil
290	6	1.5	394	1	GATR_MOUSE	P23336	mus musculu	363	6	1.5	462	1	ATPB_CHLVI	P42465	chlorobium
291	6	1.5	394	1	YBNS_YEAST	P38219	saccharomyc	364	6	1.5	462	1	CATF_MOUSE	Q94013	mus musculu
292	6	1.5	395	1	FLB2_RHIME	P13119	rhizobium m	365	6	1.5	462	1	PRYC_ECOLI	P32059	escherichia
293	6	1.5	397	1	PGK_FOUAE	O66519	equifex aeo	366	6	1.5	463	1	ENGA_MYCTU	O33212	mycobacteri
294	6	1.5	397	1	YXX7_CAEEL	Q18674	caenorhabdi	367	6	1.5	463	1	GUTA_BACSU	Q34368	bacillus su
295	6	1.5	398	1	AAAD_HUMAN	P22760	homo sapien	368	6	1.5	463	1	VNFK_ANAVA	O57302	anaeana va
296	6	1.5	398	1	PGK_RHIL0	Q98fj1	rhizobium l	369	6	1.5	465	1	TPSN_MOUSE	Q94723	mus musculu
297	6	1.5	398	1	SPEB_STRPY	P00788	streptococc	370	6	1.5	466	1	CC37_SCHPO	O94740	schizosacch
298	6	1.5	400	1	AAT_RHILP	O86459	rhicobium l	371	6	1.5	467	1	GLYD_SCHPO	O13972	schizosacch
299	6	1.5	402	1	POKM_DROME	P23757	drosophila	372	6	1.5	467	1	MMOP_ECOLI	Q47689	escherichia
300	6	1.5	403	1	NOF2_DROME	P15597	drosophila	373	6	1.5	469	1	MMO1_HORSE	Q93x25	equus cabal
301	6	1.5	403	1	TRPB_ACICA	P16706	actinobact	374	6	1.5	469	1	TBG_USTVI	P32348	usellago vi
302	6	1.5	404	1	VRHO_HUMAN	P58872	homo sapien	375	6	1.5	471	1	V8I1_TRYBB	P26326	trypanosoma
303	6	1.5	404	1	VRHO_MOUSE	P58873	mus musculu	376	6	1.5	474	1	VNFK_AZOOH	P15334	azotobacter
304	6	1.5	404	1	YK33_YEAST	P23501	saccharomyc	377	6	1.5	474	1	VNFK_AZOVI	P18856	azotobacter
305	6	1.5	405	1	THRC_METUA	Q58860	methanococc	378	6	1.5	475	1	VIBO_ECOLI	P31474	pseudomonas
306	6	1.5	409	1	LE33_ARATH	Q91mci	arabidopsis	379	6	1.5	479	1	CATY_PSEPU	Q59714	pseudomonas
307	6	1.5	410	1	PXMR_METTH	Q92628	brachydanio	380	6	1.5	480	1	MURE_RICPR	O05954	ricicetia
308	6	1.5	410	1	PSMR_METTH	Q26824	methanobact	381	6	1.5	481	1	YEAV_ECOLI	P76252	escherichia
309	6	1.5	411	1	YAL2_SCHPO	Q09685	schizosacch	382	6	1.5	482	1	NUSA_BORBU	O51740	borrella bu
310	6	1.5	412	1	BL14_NEUCR	Q92247	neutrospora	383	6	1.5	483	1	ELAV_DROME	P16914	drosophila
311	6	1.5	412	1	THI1_METAC	O8c9d4	methanosarc	384	6	1.5	483	1	KG3A_HUMAN	P45840	homo sapien
312	6	1.5	413	1	APGM_SULTO	Q97dp3	sulfolobus	385	6	1.5	484	1	CATF_HUMAN	P18265	rattus norv
313	6	1.5	413	1	IDHC_SOYBN	Q06197	glycine max	386	6	1.5	484	1	KG3A_RAT	Q9ubx1	homo sapien
314	6	1.5	414	1	APGM_SULTO	Q98a0a	hepilotobus	387	6	1.5	485	1	YV4B_CAEEL	O18411	caenorhabdi
315	6	1.5	414	1	RP54_HELPJ	Q92ic8	hepilotobus	388	6	1.5	486	1	VNS3_ROTND	P35423	human rotav
316	6	1.5	414	1	RP54_HELPJ	P56143	hepilotobus	389	6	1.5	487	1	AMPL_BOVIN	P00727	bos taurus
317	6	1.5	415	1	DCDA_PSEAE	P19572	pseudomonas	390	6	1.5	487	1	AMPL_HUMAN	P28838	homo sapien
318	6	1.5	415	1	IDHC_TOBAC	P50218	nicotiana t	391	6	1.5	487	1	AMPL_MOUSE	Q9cpd7	mus musculu
319	6	1.5	415	1	PAX2_MOUSE	P32114	mus musculu	392	6	1.5	489	1	DMH4_ECOLI	P17445	escherichia
320	6	1.5	416	1	DCDA_PSEPL	O05321	pseudomonas	393	6	1.5	490	1	MYOI_HUMAN	O15151	homo sapien
321	6	1.5	416	1	IDHC_SOLTU	P50217	solatium tub	394	6	1.5	490	1	YV01_CAEEL	O18904	caenorhabdi
322	6	1.5	416	1	PAX2_HUMAN	O02862	homo sapien	395	6	1.5	495	1	MURE_XYLPA	Q9pf85	xyella fas
323	6	1.5	416	1	SYT_TREPA	O83806	treponema p	396	6	1.5	496	1	PAL2_HUMAN	Q9upg8	homo sapien
324	6	1.5	417	1	TR16_MOUSE	Q920w1	mus musculu	397	6	1.5	498	1	YSW5_CAEEL	Q10125	caenorhabdi
325	6	1.5	419	1	GSC_DROME	P54366	drosophila	398	6	1.5	499	1	CLK2_HUMAN	P49760	homo sapien

399	6	1.5	499	1	CLK2_MOUSE	O35491	mus musculus	472	6	1.5	576	1	YIT0_YEAST	P40568	saccharomyc
400	6	1.5	501	1	COBQ_BACHD	Q9Kc10	bacillus ha	473	6	1.5	577	1	BAG3_MOUSE	Q9j1v1	mus musculus
401	6	1.5	503	1	AMPA_VIBCH	Q362w5	vibrio chol	474	6	1.5	579	1	CCBS_OENBE	Q04648	oenothera b
402	6	1.5	503	1	GGHR_CAEBL	P30635	caenorhabdi	475	6	1.5	579	1	SYF_CHLMU	Q9pk01	chlamydia m
403	6	1.5	503	1	VLI_HPV66	Q80961	human papil	476	6	1.5	580	1	CG13_YEAST	P13365	saccharomyc
404	6	1.5	503	1	YKR5_YEAST	P34240	saccharomyc	477	6	1.5	580	1	EZRI_BOVIN	P31976	bos taurus
405	6	1.5	504	1	CP7A_CRIGR	P46634	cricetulus	478	6	1.5	581	1	SYF_CHLTR	P36431	chlamydia t
406	6	1.5	504	1	FLIC_SALBU	Q06969	salmonella	479	6	1.5	583	1	YCVI_YEAST	P25639	saccharomyc
407	6	1.5	504	1	FLIC_SALDE	Q06970	salmonella	480	6	1.5	585	1	EZRI_HUMAN	P15311	homo sapien
408	6	1.5	504	1	FLIC_SALDU	Q06971	salmonella	481	6	1.5	585	1	EZRI_MOUSE	P26040	mus musculus
409	6	1.5	504	1	FLIC_SALEN	Q06972	salmonella	482	6	1.5	585	1	NEPU_THEVU	Q08751	thermoactin
410	6	1.5	504	1	FLIC_SALAC	Q06981	salmonella	483	6	1.5	588	1	PPCK_THEAC	Q9h1v2	thermoplasm
411	6	1.5	504	1	FLIC_SALRO	Q06982	salmonella	484	6	1.5	589	1	PPCK_THEVO	P58306	thermoplasm
412	6	1.5	504	1	FLIC_SALSE	Q06983	salmonella	485	6	1.5	590	1	DNAK_STRMU	O06942	streptococc
413	6	1.5	506	1	UPPE_NPVAC	P18569	autographa	486	6	1.5	593	1	Y4OA_RHISN	P55586	chizobium s
414	6	1.5	507	1	FLIC_SALBE	Q06968	salmonella	487	6	1.5	599	1	YOU3_CAEBL	P30639	caenorhabdi
415	6	1.5	507	1	MCME_RAT	O62724	rattus norv	488	6	1.5	602	1	S6AD_MOUSE	P31649	mus musculus
416	6	1.5	509	1	PSBB_GUITH	O78511	guillardia	489	6	1.5	602	1	S6AD_RAT	P31646	rattus norv
417	6	1.5	510	1	DMP1_BOVIN	Q95120	bos taurus	490	6	1.5	604	1	NUSM_ORNAN	Q36459	ornithorhyn
418	6	1.5	511	1	CP47_RABIT	P14581	oryctolagus	491	6	1.5	605	1	TC17_HUMAN	O60765	homo sapien
419	6	1.5	512	1	XNFA_CLOSR	P33558	clostridium	492	6	1.5	607	1	DNAK_STRPN	P95829	streptococc
420	6	1.5	515	1	TRPE_BACSU	P03963	bacillus su	493	6	1.5	607	1	DNAX_STRPY	P95831	streptococc
421	6	1.5	518	1	SSG1_AVEGA	P12615	avena sativ	494	6	1.5	607	1	XYNA_NEOPA	P29127	neocallima
422	6	1.5	518	1	SSG2_AVEGA	P14812	avena sativ	495	6	1.5	607	1	YSCC_YEREN	Q01244	yersinia en
423	6	1.5	518	1	THI4_SCHPO	P40386	s probabie	496	6	1.5	609	1	DNAX_STRAG	P95693	streptococc
424	6	1.5	519	1	ELAV_DROVI	P23241	drosophila	497	6	1.5	609	1	FETA_HORSE	P49066	equus cabal
425	6	1.5	520	1	HPAB_ECOLI	Q57160	escherichia	498	6	1.5	612	1	ILVD_HABIN	P14851	haemophilus
426	6	1.5	520	1	HPAB_KLEPN	Q48440	klebsiella	499	6	1.5	615	1	SECD_ECOLI	P19673	escherichia
427	6	1.5	520	1	RP54_RH1ET	P49989	rhizobium e	500	6	1.5	615	1	SECD_SALCH	Q9zff8	salmonella
428	6	1.5	522	1	LNT_HAFIN	P44626	haemophilus	501	6	1.5	617	1	ILVD_CAUCR	P55186	caulobacter
429	6	1.5	522	1	POLS_RUBV	P08564	rubella vir	502	6	1.5	617	1	SPH2_MOUSE	Q9j1a7	mus musculus
430	6	1.5	526	1	CATA_CAVPO	Q64405	cavia porce	503	6	1.5	619	1	DNAX_CLOPE	P26823	clostridium
431	6	1.5	527	1	CATA_RANRU	Q9pwf7	rana rugosa	504	6	1.5	619	1	ILVD_NEIMA	Q9jue0	neisseria m
432	6	1.5	530	1	UTR1_YEAST	P21373	saccharomyc	505	6	1.5	619	1	ILVD_NEIMB	Q9j661	neisseria m
433	6	1.5	532	1	DCTS_BACHD	Q9k997	bacillus ha	506	6	1.5	619	1	YA16_CHLNP	Q9z6p3	chlamydia p
434	6	1.5	532	1	YOAD_ECOLI	P76261	escherichia	507	6	1.5	626	1	CC23_YEAST	P16522	saccharomyc
435	6	1.5	534	1	VLI_HPV56	P36743	human papil	508	6	1.5	634	1	HNFA_HUMAN	Q90867	gallus gall
436	6	1.5	536	1	ENTE_ECO57	Q8xbv3	escherichia	509	6	1.5	638	1	LIK2_HUMAN	P53671	homo sapien
437	6	1.5	536	1	ENTE_ECOLI	P10378	escherichia	510	6	1.5	638	1	LIK2_RAT	P53666	rattus norv
438	6	1.5	536	1	YRN3_CAEBL	Q09606	caenorhabdi	511	6	1.5	642	1	LIK2_CHICK	Q9z577	pyrococcus
439	6	1.5	538	1	IPGD_SHIFL	Q07566	shigella fl	512	6	1.5	648	1	TORA_PYPAB	O9z577	pyrococcus
440	6	1.5	538	1	IPGD_SHISO	Q55286	shigella so	513	6	1.5	648	1	TORA_PYPHO	O58495	pyrococcus
441	6	1.5	539	1	MS33_HYDAT	P33922	hydra atten	514	6	1.5	649	1	HS73_YEAST	P09435	saccharomyc
442	6	1.5	539	1	VGLF_P13H4	P06828	human parai	515	6	1.5	650	1	HS71_LYCES	P24629	lycopersico
443	6	1.5	540	1	KLC_CAEBL	P46822	caenorhabdi	516	6	1.5	652	1	NOS2_PARDE	Q51705	paracoccus
444	6	1.5	541	1	VIBE_HELPY	P56128	helicobacte	517	6	1.5	654	1	SPH2_HUMAN	Q9nra0	homo sapien
445	6	1.5	543	1	VIBE_VIBCH	O07899	vibrio chol	518	6	1.5	655	1	SKM1_YEAST	Q12469	saccharomyc
446	6	1.5	544	1	GSHC_SOVEN	P48640	glycine max	519	6	1.5	656	1	ACSA_RHOCA	O68040	rhodobacter
447	6	1.5	547	1	BK42_DROME	P35736	drosophila	520	6	1.5	658	1	RAE2_HUMAN	P26374	homo sapien
448	6	1.5	551	1	RM56_MOUSE	Q9ep89	mus musculus	521	6	1.5	658	1	SOHC_ZYMMO	P33990	zymomonas m
449	6	1.5	553	1	GGPI_YEAST	P12709	saccharomyc	522	6	1.5	661	1	DHSA_DROME	Q94523	drosophila
450	6	1.5	554	1	CA1A_PARDE	P08305	paracoccus	523	6	1.5	661	1	WHI3_YEAST	P34761	saccharomyc
451	6	1.5	556	1	SVR_BACHD	Q9k6c1	bacillus ha	524	6	1.5	664	1	YFGK_CAEBL	O18696	caenorhabdi
452	6	1.5	556	1	YDP4_SCHPO	O14013	schizosacch	525	6	1.5	676	1	RNR_CHLNP	Q9z848	chlamydia p
453	6	1.5	557	1	GGPI_KLUUA	P12341	kluveromyc	526	6	1.5	684	1	TC10_YEAST	P50273	saccharomyc
454	6	1.5	557	1	ILVB_BACSU	P51785	bacillus su	527	6	1.5	685	1	ATKB_CLOAB	Q32328	clostridium
455	6	1.5	562	1	ILVB_ECOLI	P08142	escherichia	528	6	1.5	686	1	KITH_HSVBM	P33802	bovine herp
456	6	1.5	564	1	ATR_HUMAN	Q9h6x2	homo sapien	529	6	1.5	686	1	MEPD_RAT	P24155	rattus norv
457	6	1.5	564	1	YEDQ_ECO57	Q8xb92	escherichia	530	6	1.5	690	1	CVG3_RAT	P19686	rattus norv
458	6	1.5	564	1	YEDQ_ECOLI	P76330	escherichia	531	6	1.5	697	1	CEAD_ECOLI	P17998	escherichia
459	6	1.5	566	1	HENA_IADIR	P11132	influenza a	532	6	1.5	702	1	FOXA_SALTY	Q56145	salmonella
460	6	1.5	566	1	HENA_IATKI	P11135	influenza a	533	6	1.5	703	1	DD35_HUMAN	Q9h5z1	homo sapien
461	6	1.5	569	1	S6AD_HUMAN	Q9nads	homo sapien	534	6	1.5	705	1	PAL2_LITER	O49836	lithospermu
462	6	1.5	569	1	Y397_MYCPN	Q50333	mycoplasma	535	6	1.5	706	1	ARN2_HUMAN	Q9hbz2	homo sapien
463	6	1.5	572	1	PTI_BACHD	Q9k8d3	bacillus ha	536	6	1.5	709	1	FTSI_NEPOL	Q9t136	nephroselm
464	6	1.5	572	1	PTI_LISIN	Q92d19	listeria in	537	6	1.5	709	1	SYT_ARATH	Q04630	arabidopsis
465	6	1.5	572	1	PTI_LISMO	O31149	listeria mo	538	6	1.5	712	1	ARN2_MOUSE	Q61324	mus musculus
466	6	1.5	572	1	PTI_STAAN	Q931u2	staphylococ	539	6	1.5	714	1	FXP2_MOUSE	P58463	mus musculus
467	6	1.5	572	1	PTI_STPAU	Q95v14	staphylococ	540	6	1.5	715	1	PAB3_STRGR	O15409	homo sapien
468	6	1.5	572	1	PTI_MYCCA	P51183	staphylococ	541	6	1.5	723	1	TAP1_RAT	P32483	streptomyce
469	6	1.5	573	1	PTI_MYCCA	P45617	mycoplasma	542	6	1.5	725	1	TAP1_RAT	P36370	rattus norv
470	6	1.5	574	1	PTI_SPACA	P23533	staphylococ	543	6	1.5	730	1	CATA_HALWA	O59651	haloarcu
471	6	1.5	575	1	MIS_BOVIN	P03972	bos taurus	544	6	1.5	731	1	P13A_HUMAN	P00488	homo sapien

545	6	1.5	731	1	NGP1_HUMAN	013823 homo sapien	618	6	1.5	900	1	SYA_MYCSE	P47534 mycoplasma
546	6	1.5	732	1	DD35_CAEEL	Q9bqx8 caenorhabdi	619	6	1.5	903	1	DPOL_BPR69	Q36087 bacterioph
547	6	1.5	734	1	PSAB_CHLYU	P56342 chlorella v	620	6	1.5	908	1	VP2_BRD	P35934 broadhaven
548	6	1.5	734	1	PSAB_NEPOL	Q9tkw1 neptroleini	621	6	1.5	911	1	B3AT_HUMAN	P02730 homo sapien
549	6	1.5	734	1	PSAB_PORPU	P51285 porphyra pu	622	6	1.5	922	1	B3AT_CHICK	P15575 gallus galli
550	6	1.5	735	1	DD15_SCHPO	042245 schizosacch	623	6	1.5	926	1	MAV3_SCHCO	P37934 schizophy11
551	6	1.5	735	1	PSAB_CHLMO	P36492 chlamydomon	624	6	1.5	927	1	B3AT_RAT	P23562 rattus norv
552	6	1.5	735	1	PSAB_CHLRE	P09144 chlamydomon	625	6	1.5	928	1	VGIB_MCWVS	P27171 murine cyto
553	6	1.5	738	1	PSAB_SYNPM	Q9rcfc synchococc	626	6	1.5	929	1	B3AT_MOUSE	P04491 mus musculu
554	6	1.5	739	1	DD15_CAEEL	Q20875 caenorhabdi	627	6	1.5	933	1	SLAP_CAMEP	P35827 campylobact
555	6	1.5	741	1	YD35_MYCPN	P75443 mycoplasma	628	6	1.5	960	1	FGD1_MOUSE	P52734 mus musculu
556	6	1.5	742	1	DBP7_YEAST	P36120 saccharomyc	629	6	1.5	961	1	BASO_MOUSE	Q35914 mus musculu
557	6	1.5	742	1	PSAB_PROS1	Q9rc07 prochloroto	630	6	1.5	961	1	FGD1_HUMAN	P98174 homo sapien
558	6	1.5	744	1	RELA_VIBSS	P55133 vibrio sp.	631	6	1.5	967	1	ATSI_RAT	Q9wuy1 rattus norv
559	6	1.5	747	1	PSAB_PROMA	Q9rcv0 prochloroto	632	6	1.5	968	1	ATSI_MOUSE	P97857 mus musculu
560	6	1.5	752	1	FOH1_MOUSE	Q35409 m glutamate	633	6	1.5	974	1	YMB4_CAEEL	Q03601 caenorhabdi
561	6	1.5	752	1	PARC_SALTY	P26973 salmonella	634	6	1.5	976	1	AMY_BUTPI	P30269 butyrivibri
562	6	1.5	754	1	YAJ3_SCHPO	Q09903 schizosacch	635	6	1.5	984	1	NOF_DROME	P16320 dirosophila
563	6	1.5	755	1	NAP4_DESDE	P81186 desulfovibr	636	6	1.5	990	1	YAA2_SCHPO	Q09796 schizosacch
564	6	1.5	757	1	ECR_LOCCU	Q18531 lucilia cup	637	6	1.5	991	1	BMP1_MOUSE	P98063 mus musculu
565	6	1.5	758	1	DD15_MOUSE	Q35286 mus musculu	638	6	1.5	992	1	POLS_RUBVM	P08563 rubella vir
566	6	1.5	758	1	SP21_YEAST	P35209 saccharomyc	639	6	1.5	993	1	POLS_IBDVP	P25220 avian infec
567	6	1.5	759	1	GSPD_XANCP	P29941 xanthomonas	640	6	1.5	994	1	BASO_HUMAN	Q01954 homo sapien
568	6	1.5	760	1	PXN1_MOUSE	Q08696 mus musculu	641	6	1.5	1008	1	MOG4_CAEEL	Q45244 caenorhabdi
569	6	1.5	760	1	RELA_CORGL	Q08731 corynebacte	642	6	1.5	1011	1	VPS3_YEAST	P23643 saccharomyc
570	6	1.5	763	1	DYRA_HUMAN	Q13627 homo sapien	643	6	1.5	1012	1	POLG_IBDVO	P27276 avian infec
571	6	1.5	763	1	DYRA_MOUSE	Q61214 mus musculu	644	6	1.5	1012	1	POLS_IBDVS	P25219 avian infec
572	6	1.5	763	1	DYRA_RAT	Q63470 rattus norv	645	6	1.5	1012	1	POLS_IBDVA	P08364 avian infec
573	6	1.5	767	1	PR43_YEAST	P53131 saccharomyc	646	6	1.5	1012	1	POLS_IBDVC	P15480 avian infec
574	6	1.5	772	1	GEUS_PLG	P23035 sus scrofa	647	6	1.5	1012	1	POLS_IBDVS	P22351 avian infec
575	6	1.5	775	1	AD28_HUMAN	Q9ukq2 homo sapien	648	6	1.5	1018	1	HMV1_MYCPN	Q50365 mycoplasma
576	6	1.5	790	1	AD30_HUMAN	Q9ukf2 homo sapien	649	6	1.5	1018	1	YDIJ_ECOLI	P77748 escherichia
577	6	1.5	793	1	PHK2_ANASP	Q8YVZ6 anabaena sp	650	6	1.5	1023	1	RT12_ACTPL	P55128 actinobacil
578	6	1.5	794	1	TRK1_LYMAST	Q76997 lymanaea str	651	6	1.5	1023	1	RT12_ACTPL	P55128 actinobacil
579	6	1.5	795	1	DD15_HUMAN	Q43143 homo sapien	652	6	1.5	1025	1	DEYD_BOVIN	Q28007 bos taurus
580	6	1.5	796	1	ABPA_EMENT	P20345 emericella	653	6	1.5	1025	1	DEYD_HUMAN	Q12882 homo sapien
581	6	1.5	805	1	DF19_CAEEL	Q09555 caenorhabdi	654	6	1.5	1034	1	BGAL_BACME	Q52847 bacillus me
582	6	1.5	805	1	VNCS_ADEDB	Q90185 aedes albop	655	6	1.5	1034	1	GCSB_FLAAR	P43362 flaveria pr
583	6	1.5	806	1	SYL_BACHD	Q9K768 bacillus ha	656	6	1.5	1034	1	GCSB_FLAAR	Q49850 flaveria pr
584	6	1.5	807	1	PHK_RHILLO	Q988V7 rhizobium l	657	6	1.5	1034	1	GCSF_FLAAR	Q49852 flaveria tr
585	6	1.5	808	1	YNPB_CAEEL	P34561 caenorhabdi	658	6	1.5	1035	1	GCSF_SOLTR	Q49954 botanum tub
586	6	1.5	809	1	AK2H_ECOLI	P00562 escherichia	659	6	1.5	1037	1	GCSA_FLAAR	P43361 flaveria pr
587	6	1.5	814	1	SLAI_BACAN	P49051 bacillus an	660	6	1.5	1041	1	BGT2_YEAST	P42835 saccharomyc
588	6	1.5	815	1	AKH_HABIN	P44505 haemophilus	661	6	1.5	1044	1	GCSF_ARATH	Q80988 arabidopsis
589	6	1.5	821	1	MC06_HUMAN	Q14566 homo sapien	662	6	1.5	1052	1	BU1B_MOUSE	Q92120 mus musculu
590	6	1.5	821	1	MC06_MOUSE	P97311 mus musculu	663	6	1.5	1054	1	R0GB_DROME	P43125 drosophila
591	6	1.5	822	1	DMSA_RHOSH	Q57366 rhodobacter	664	6	1.5	1057	1	GCSF_PEA	P26969 plium sativ
592	6	1.5	826	1	HIFR_HUMAN	Q16665 homo sapien	665	6	1.5	1057	1	RPOC_STAAU	P47770 staphylococ
593	6	1.5	827	1	6P21_YEAST	P40433 saccharomyc	666	6	1.5	1057	1	TLD_DROME	P25723 drosophila
594	6	1.5	832	1	ALP6_SCHPO	Q9U8Q2 schizosacch	667	6	1.5	1063	1	POLS_RUBVA	P21480 rubella vir
595	6	1.5	834	1	CASL_HUMAN	Q14511 homo sapien	668	6	1.5	1063	1	POLS_RUBVR	P19725 rubella vir
596	6	1.5	834	1	DPOT_THREH	P52028 thermus the	669	6	1.5	1063	1	POLS_RUBVT	P07566 rubella vir
597	6	1.5	845	1	RIR1_TREPA	Q83372 treponema p	670	6	1.5	1064	1	CARB_IACAC	Q13371 lactococcus
598	6	1.5	852	1	SRCH_RABIT	P16230 oryctolagus	671	6	1.5	1070	1	PKT7_HUMAN	Q13308 homo sapien
599	6	1.5	860	1	YERK_SCHPO	Q36021 schizosacch	672	6	1.5	1089	1	MM1B_MYCTU	Q07800 mycobacteri
600	6	1.5	861	1	SYI_MYCPN	P75258 mycoplasma	673	6	1.5	1089	1	NMD2_YEAST	P38798 saccharomyc
601	6	1.5	862	1	MUTS_BORBU	Q51737 borrelia bu	674	6	1.5	1097	1	EX5C_MYCTU	P46591 mycobacteri
602	6	1.5	862	1	PGCV_MACNE	Q28858 macaca neme	675	6	1.5	1117	1	CTT4_INBUR	P47950 neurospora
603	6	1.5	863	1	AR56_YEAST	Q01417 s argys, 6 pr	676	6	1.5	1121	1	DDX8_ARATH	Q36953 arabidopsis
604	6	1.5	863	1	S1P1_YEAST	P32578 saccharomyc	677	6	1.5	1126	1	DNBI_HSVGA	P24910 herpesvirs
605	6	1.5	871	1	UL47_HSVB	P28929 equine herp	678	6	1.5	1138	1	PHY_PICAB	Q40762 plectra ables
606	6	1.5	872	1	UL47_HSV4	P25073 equine herp	679	6	1.5	1144	1	MZAI_HUMAN	Q16706 homo sapien
607	6	1.5	874	1	SYA_HABIN	P43815 haemophilus	680	6	1.5	1145	1	DPG1_DROME	Q27607 drosophila
608	6	1.5	874	1	SYA_PASMU	P57933 paeteurella	681	6	1.5	1162	1	LEPR_RAT	Q62959 rattus norv
609	6	1.5	880	1	BGL2_SACPT	P22307 saccharomyc	682	6	1.5	1168	1	DDX8_SCHPO	Q42643 schizosacch
610	6	1.5	884	1	IF2_YERPE	Q820C2 yerinia pe	683	6	1.5	1173	1	ATC2_YEAST	P38929 saccharomyc
611	6	1.5	885	1	LOM1_MAIZE	P93447 zea mays (m	684	6	1.5	1181	1	YU02_METVA	Q63031 methanococc
612	6	1.5	886	1	YEB6_SCHPO	Q14302 schizosacch	685	6	1.5	1214	1	TGSA_RAT	Q63079 rattus norv
613	6	1.5	890	1	KPCN_HUMAN	Q94806 homo sapien	686	6	1.5	1220	1	DDX8_HUMAN	Q14562 homo sapien
614	6	1.5	893	1	DAG1_MOUSE	Q62165 mus musculu	687	6	1.5	1227	1	B3A3_MOUSE	P16283 mus musculu
615	6	1.5	894	1	WPR4_BACSU	P54423 bacillus su	688	6	1.5	1227	1	B3A3_RAT	P23283 rattus norv
616	6	1.5	898	1	C1Z1_HUMAN	Q9ULV3 homo sapien	689	6	1.5	1232	1	B3A3_HUMAN	P48751 homo sapien
617	6	1.5	898	1	IF38_CAEEL	Q02328 caenorhabdi	690	6	1.5	1233	1	B3A3_FABIT	Q18917 oryctolagus

691 6 1.5 1234 1 B3A2 RAT  
692 6 1.5 1235 1 TRK1 YEAST  
693 6 1.5 1237 1 B3A2 MOUSE  
694 6 1.5 1237 1 B3A2 RABBIT  
695 6 1.5 1238 1 B3A2 CAVO  
696 6 1.5 1241 1 B3A2 HUMAN  
697 6 1.5 1241 1 PER DROPS  
698 6 1.5 1251 1 Y0U3 CAEL  
699 6 1.5 1263 1 DP2L PYRFU  
700 6 1.5 1283 1 PEX1 HUMAN  
701 6 1.5 1290 1 BXC1 CLOSTRIDIUM  
702 6 1.5 1294 1 RPO WCMVM  
703 6 1.5 1294 1 RPO WCMVO  
704 6 1.5 1295 1 BXA1 CLOBO  
705 6 1.5 1312 1 PUTA KLEAE  
706 6 1.5 1317 1 RPOD SYN3  
707 6 1.5 1320 1 PUTA ECOLI  
708 6 1.5 1320 1 PUTA SALTY  
709 6 1.5 1331 1 MANB CALSA  
710 6 1.5 1365 1 GTFS STRDO  
711 6 1.5 1365 1 SUZ2 DROME  
712 6 1.5 1381 1 YBET YEAST  
713 6 1.5 1437 1 VGLM BUNGE  
714 6 1.5 1448 1 UGG SCHPO  
715 6 1.5 1468 1 ROGB AQUAE  
716 6 1.5 1487 1 BLM DROME  
717 6 1.5 1548 1 UGG DROME  
718 6 1.5 1555 1 GDE RABIT  
719 6 1.5 1562 1 RPOD CHLVU  
720 6 1.5 1584 1 U104 CAEL  
721 6 1.5 1627 1 PAPA HUMAN  
722 6 1.5 1629 1 GLS2 ARATH  
723 6 1.5 1645 1 OMPB RICTY  
724 6 1.5 1654 1 PCFB HUMAN  
725 6 1.5 1744 1 TANA XENIA  
726 6 1.5 1802 1 HKR1 YEAST  
727 6 1.5 1829 1 DPOL THEST  
728 6 1.5 1856 1 MGA HUMAN  
729 6 1.5 1857 1 FAS2 PENPA  
730 6 1.5 1872 1 T2D1 HUMAN  
731 6 1.5 1922 1 TRIB HUMAN  
732 6 1.5 1997 1 OTOP HUMAN  
733 6 1.5 1997 1 OTOP HUMAN  
734 6 1.5 2144 1 GLUT1 YEAST  
735 6 1.5 2280 1 COAC SCHPO  
736 6 1.5 2334 1 WAPA BACSU  
737 6 1.5 2481 1 UN52 CAEL  
738 6 1.5 2504 1 FAS HUMAN  
739 6 1.5 2549 1 FRAP HUMAN  
740 6 1.5 2549 1 FRAP MOUSE  
741 6 1.5 2549 1 FRAP RAT  
742 6 1.5 2554 1 7LES DROME  
743 6 1.5 2663 1 CENE HUMAN  
744 6 1.5 2907 1 FEN2 MOUSE  
745 6 1.5 2911 1 FEN2 HUMAN  
746 6 1.5 3014 1 CUR1 HUMAN  
747 6 1.5 3122 1 DPO2 MOUSE  
748 6 1.5 3137 1 CA36 CHICK  
749 6 1.5 3175 1 RPOA EAV  
750 6 1.5 3178 1 YS89 CAEL  
751 6 1.5 3396 1 PGCV HUMAN  
752 6 1.5 3433 1 UTR0 HUMAN  
753 6 1.5 4036 1 RRPL DUGBV  
754 6 1.5 4085 1 RRP1 CVH22  
755 6 1.5 4303 1 PKD1 HUMAN  
756 6 1.5 4568 1 DHVC CAEL  
757 6 1.5 4644 1 FAT DROME  
758 6 1.5 5147 1 FAT DROME  
759 6 1.5 5179 1 MUC2 HUMAN  
760 6 1.5 6486 1 TYCC BACBR  
761 6 1.5 15 1 TRPA LEUMA  
762 5 1.2 15 1 H5\_COTJA  
763 5 1.2 16 1

P23347 rattus norv  
P12685 saccharomyc  
P13808 mus musculu  
P48746 oryctolagus  
P92088 cavia porce  
P04920 homo sapien  
P12348 drosophila  
Q09550 caenorhabdi  
P81409 pyrococcus  
P13933 homo sapien  
P18640 clostridium  
P09498 white clove  
P15402 white clove  
P10845 clostridium  
Q45894 clostridium  
O52485 klebsiella  
P73334 synechocyst  
P09546 escherichia  
P10503 salmonella  
P22533 caldococcus  
P29336 streptococc  
P25172 drosophila  
P34216 saccharomyc  
P12430 bunyavirus  
Q09140 schizosacch  
O67762 aquifex ae  
Q9vg18 drosophila  
Q09332 drosophila  
P35574 o glycogen  
P12465 chlorella v  
P23678 caenorhabdi  
Q13219 homo sapien  
P910P4 arabidopsis  
P69689 r outer mem  
Q94913 homo sapien  
P01550 xenopus lae  
P41809 saccharomyc  
O33845 thermococcu  
O43451 homo sapien  
P15368 p fatty aci  
P21675 homo sapien  
Q14669 homo sapien  
O9hc10 homo sapien  
Q986f1 mus musculu  
Q12680 saccharomyc  
P78820 schizosacch  
Q07833 bacillus su  
O06561 caenorhabdi  
P49327 homo sapien  
P42345 homo sapien  
Q9j1n9 mus musculu  
P42346 rattus norv  
P13368 drosophila  
Q02224 homo sapien  
P61555 mus musculu  
P35556 mus sapien  
Q9dyq6 homo sapien  
P61493 mus musculu  
P15989 gallus gall  
P19811 equine arte  
Q09624 caenorhabdi  
P13611 homo sapien  
P46939 homo sapien  
Q66431 dugbe virus  
Q05002 human coron  
P98161 homo sapien  
Q19020 caenorhabdi  
Q9jhu4 mus musculu  
P33450 drosophila  
Q02817 homo sapien  
Q30409 b tyrocidin  
P81753 leucophaea  
P18638 coturnix co

1 17 1.2 5 764  
1 19 1.2 5 765  
1 20 1.2 5 766  
1 20 1.2 5 767  
1 20 1.2 5 768  
1 20 1.2 5 769  
1 32 1.2 5 770  
1 33 1.2 5 771  
1 33 1.2 5 772  
1 34 1.2 5 773  
1 37 1.2 5 774  
1 37 1.2 5 775  
1 37 1.2 5 776  
1 38 1.2 5 777  
1 38 1.2 5 778  
1 38 1.2 5 779  
1 38 1.2 5 780  
1 38 1.2 5 781  
1 38 1.2 5 782  
1 38 1.2 5 783  
1 38 1.2 5 784  
1 40 1.2 5 785  
1 43 1.2 5 786  
1 44 1.2 5 787  
1 48 1.2 5 788  
1 48 1.2 5 789  
1 50 1.2 5 790  
1 52 1.2 5 791  
1 52 1.2 5 792  
1 53 1.2 5 793  
1 54 1.2 5 794  
1 55 1.2 5 795  
1 56 1.2 5 796  
1 58 1.2 5 797  
1 58 1.2 5 798  
1 59 1.2 5 799  
1 60 1.2 5 800  
1 60 1.2 5 801  
1 60 1.2 5 802  
1 60 1.2 5 803  
1 60 1.2 5 804  
1 60 1.2 5 805  
1 60 1.2 5 806  
1 60 1.2 5 807  
1 60 1.2 5 808  
1 61 1.2 5 809  
1 61 1.2 5 810  
1 61 1.2 5 811  
1 62 1.2 5 812  
1 62 1.2 5 813  
1 62 1.2 5 814  
1 62 1.2 5 815  
1 63 1.2 5 816  
1 63 1.2 5 817  
1 64 1.2 5 818  
1 64 1.2 5 819  
1 65 1.2 5 820  
1 65 1.2 5 821  
1 66 1.2 5 822  
1 66 1.2 5 823  
1 67 1.2 5 824  
1 67 1.2 5 825  
1 67 1.2 5 826  
1 67 1.2 5 827  
1 67 1.2 5 828  
1 67 1.2 5 829  
1 67 1.2 5 830  
1 68 1.2 5 831  
1 68 1.2 5 832  
1 69 1.2 5 833  
1 69 1.2 5 834  
1 69 1.2 5 835  
1 70 1.2 5 836

P80069 mycobacteri  
P14539 capirus ter  
P80974 thunnus obe  
P80329 oncorhynchu  
P80928 sus scrofa  
O78512 guillardia  
P32424 rana brevip  
P80330 oncorhynchu  
P40839 rana esculu  
P12260 bos taurus  
P29301 arabidopsis  
P12166 oryza sativ  
P46306 chlamydomon  
P22974 chlamydomon  
P56339 chlorella v  
Q9cm22 cyanidium c  
P12228 euglena gra  
O78464 guillardia  
P12165 marchantia  
P49544 odoncella s  
P51389 porphyra pu  
P21376 melegria g  
P20312 bacterioph  
P80569 carnobacter  
P80573 alicigenes  
Q9ppn5 ureaplasma  
Q29329 sus scrofa  
Q02513 human papil  
P56507 haemophilus  
O26357 methanobact  
P33710 canis famil  
Q06173 desulfovibr  
P10946 bacillus su  
P16961 streptomyce  
Q8yvm0 anabaena sp  
P35087 synechococc  
P06563 bacillus su  
P32879 alpyaurus l  
P01438 aestrotia st  
P25492 enhydrina s  
P25494 hydrophis c  
P01437 hydrophis l  
P25493 enhydrina s  
P19958 alpyaurus l  
P25683 cendroaspi  
P17591 human adeno  
O54457 erwina her  
P09401 methanococc  
P09407 momordica c  
P01434 acanthophis  
P25497 pseudochis  
P34521 caenorhabdi  
P46162 bos taurus  
P19279 thermoprote  
P19279 bos taurus  
P30390 euglena gra  
P30773 bacterioph  
P95070 mycobacteri  
O78517 guillardia  
O78517 oviss aries  
Q04742 mus musculu  
O99598 artibeus ja  
P38592 halichoerus  
P49499 microtus pe  
Q00522 phoca vitul  
Q9mev6 sciurus vul  
Q9mb2 talpa europ  
P56053 helicobacte  
Q9zyy7 hippopotamu  
Q9prj9 pseudopleur  
Q47840 enterococcu  
P31969 synechococc  
P31567 galdieria s  
P05115 lycoperisico

837	5	1.2	71	1	IF1_LACIA	P2749	lactococcus	910	5	1.2	92	1	RPOL_PYRAE	Q8246	pyrobaculum
838	5	1.2	71	1	LIT3_CABEL	Q10574	caenorhabdi	911	5	1.2	92	1	SFSB_ECOLI	P18837	escherichia
839	5	1.2	71	1	NU4M_ARTSA	P19046	artemia sal	912	5	1.2	93	1	GATC_HELPU	Q92486	helicobacte
840	5	1.2	71	1	YC40_PORPU	P51344	porphyra pu	913	5	1.2	93	1	GATC_HELPU	O25622	helicobacte
841	5	1.2	72	1	IF1_STRPN	Q97800	streptococc	914	5	1.2	93	1	NHPA_YEAST	P11332	saccharomyc
842	5	1.2	72	1	IF1_STRPY	Q9A173	streptococc	915	5	1.2	93	1	YF83_YEAST	P14664	saccharomyc
843	5	1.2	73	1	H171_HUMAN	Q43727	homo sapien	916	5	1.2	93	1	YQ32_MYCTU	P71931	mycobacteri
844	5	1.2	73	1	HFO_BACSU	Q31796	bacillus su	917	5	1.2	94	1	SABE_SARPE	P18133	sarcopaga
845	5	1.2	73	1	SAS5_BACME	P41459	autographa	918	5	1.2	94	1	SCRV_VIBAL	P24508	vibrio algi
846	5	1.2	73	1	Y055_NPVAC	P04835	autographa	919	5	1.2	94	1	V45_BPT3	P20318	bacteriopia
847	5	1.2	74	1	HOLI_BBDP1	P49542	odontella s	920	5	1.2	94	1	YRHB_ECOLI	P46857	escherichia
848	5	1.2	74	1	YCA7_ODOI1	P31280	bacteriopia	921	5	1.2	95	1	GLRX_RICPR	Q92426	thermus the
849	5	1.2	75	1	VGB_BPAL3	Q10781	mycobacteri	922	5	1.2	95	1	NOOB_THERT	O56226	thermus the
850	5	1.2	75	1	YF45_MYCTU	Q07490	rattus norv	923	5	1.2	95	1	SR19_PYRAE	Q92478	pyrobaculum
851	5	1.2	76	1	CHAB_ECOLI	P39162	escherichia	924	5	1.2	95	1	Y060_NPVAD	O90178	lymantria d
852	5	1.2	76	1	SECG_BACSU	Q32233	bacillus su	925	5	1.2	95	1	YF55_ARCTU	O28717	archaeoglob
853	5	1.2	76	1	TX3A_BACAP	P33034	agelaeopsis	926	5	1.2	96	1	EXOX_RHISN	P14185	rhizobium s
854	5	1.2	76	1	VG06_BPMU	Q38477	bacteriopia	927	5	1.2	96	1	GATC_ANASP	P58527	anabaena sp
855	5	1.2	77	1	MCPI_LYCES	P01076	lycopersico	928	5	1.2	96	1	GATC_DEIRA	Q92403	deinococcus
856	5	1.2	77	1	Y4D2_RHISN	P54409	rhizobium s	929	5	1.2	96	1	RL37_HUMAN	O43020	schizosacch
857	5	1.2	78	1	GV42_HALNI	P08959	halobacteri	930	5	1.2	96	1	Y448_NQDSE	O66756	aquifex ae
858	5	1.2	78	1	HFO_BACPD	Q9KAC4	bacillus ha	931	5	1.2	96	1	YDRX_ECOLI	P76165	escherichia
859	5	1.2	79	1	DLIC_STRPY	P17307	bacteriopia	932	5	1.2	96	1	CSQA_THINE	P45689	thiobacilli
860	5	1.2	79	1	RL22_PRUPA	Q44161	prunus arme	933	5	1.2	97	1	NHP1_BABBO	P40632	babesia bov
861	5	1.2	79	1	U197_CAEEL	Q95611	caenorhabdi	934	5	1.2	97	1	VG1C_BPPZA	P06949	bacteriopia
862	5	1.2	80	1	ATPD_PIG	Q95312	sus scrofa	935	5	1.2	98	1	C67A_LOCMT	P80031	locusta mig
863	5	1.2	80	1	PSBE_SYNY3	P09190	synecocyst	936	5	1.2	98	1	CME2_BACSU	P39146	bacillus su
864	5	1.2	80	1	NUEM_NANPL	Q06059	anae plactyr	937	5	1.2	98	1	CMEC_BACSU	P25955	bacillus su
865	5	1.2	81	1	NXS2_ALPLA	P19959	alpyrusus 1	938	5	1.2	98	1	COXB_BOVIN	P00428	boe taurus
866	5	1.2	81	1	POC3_SYRVU	P19960	alpyrusus 1	939	5	1.2	98	1	ELIB_PHYDM	P35697	phytochlor
867	5	1.2	81	1	RUXG_MEDSA	P58171	styringia vul	940	5	1.2	98	1	FER2_PHYM	P00031	phytolacca
868	5	1.2	81	1	RUXG_MEDSA	P24715	medicago sa	941	5	1.2	98	1	S113_BOVIN	P79342	boe taurus
869	5	1.2	82	1	HYBG_ECOLI	P31785	escherichia	942	5	1.2	98	1	S113_BOVIN	Q95984	homo sapien
870	5	1.2	82	1	P217_MERUN	Q9P963	merionesa un	943	5	1.2	98	1	Y095_MYCTU	Q10891	mycobacteri
871	5	1.2	82	1	Y024_RICPR	Q92600	ricetetsia	944	5	1.2	98	1	CYT_NAJAT	P81714	naja atra (
872	5	1.2	82	1	CALD_BOVIN	Q27976	boe taurus	945	5	1.2	99	1	GATC_RALSO	O87322	raistonia s
873	5	1.2	83	1	CLSG_HAEMA	Q25088	haemopsis ma	946	5	1.2	99	1	PLAS_CAPRU	P00094	caprella bu
874	5	1.2	83	1	POC1_ARATH	Q9SRP7	arabidopsis	947	5	1.2	99	1	SR19_PYRHO	O74102	pyrococcus
875	5	1.2	83	1	POC2_BRARA	Q39406	brassica ra	948	5	1.2	99	1	VH8B_BPT7	P03751	bacteriopia
876	5	1.2	83	1	PSB8_CHLUV	P56309	chlorrella v	949	5	1.2	99	1	YAN4_YEAST	P39563	saccharomyc
877	5	1.2	83	1	Y0GB_ECOLI	P46879	escherichia	950	5	1.2	99	1	YCUE_SALTY	P16656	salmoneilla
878	5	1.2	84	1	POC3_OLEBU	O81092	olea europoa	951	5	1.2	99	1	PLAS_CAPRU	O57423	escherichia
879	5	1.2	84	1	Y149_ARCFU	Q30088	archaeoglob	952	5	1.2	100	1	RR14_EPIVY	P27071	epifagus vi
880	5	1.2	84	1	NUMM_ARTSP	Q37751	artemia san	953	5	1.2	100	1	URE3_MYCTU	P50043	mycobacteri
881	5	1.2	85	1	POC4_ALNGL	O81701	alnus gluct	954	5	1.2	100	1	DMT1_ALUMI	Q9Pue0	alligator m
882	5	1.2	85	1	POC4_BETVE	Q39419	betula verr	955	5	1.2	101	1	NUIC_ORYSA	P12128	oryza sativ
883	5	1.2	85	1	RMJ3_YEAST	P20084	saccharomyc	956	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
884	5	1.2	85	1	YRDB_ECOLI	P45795	escherichia	957	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
885	5	1.2	85	1	NUMM_ARTSA	P19049	artemia sal	958	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
886	5	1.2	86	1	NUMM_ARTSA	P54899	artemia sal	959	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
887	5	1.2	86	1	YRDB_ECOLI	P38455	saccharomyc	960	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
888	5	1.2	86	1	YRDB_ECOLI	P38455	saccharomyc	961	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
889	5	1.2	86	1	YRDB_ECOLI	P38455	saccharomyc	962	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
890	5	1.2	86	1	YRDB_ECOLI	P38455	saccharomyc	963	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
891	5	1.2	86	1	YRDB_ECOLI	P38455	saccharomyc	964	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
892	5	1.2	86	1	YRDB_ECOLI	P38455	saccharomyc	965	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
893	5	1.2	87	1	YRDB_ECOLI	P38455	saccharomyc	966	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
894	5	1.2	87	1	YRDB_ECOLI	P38455	saccharomyc	967	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
895	5	1.2	87	1	YRDB_ECOLI	P38455	saccharomyc	968	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
896	5	1.2	87	1	YRDB_ECOLI	P38455	saccharomyc	969	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
897	5	1.2	88	1	YRDB_ECOLI	P38455	saccharomyc	970	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
898	5	1.2	88	1	YRDB_ECOLI	P38455	saccharomyc	971	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
899	5	1.2	88	1	YRDB_ECOLI	P38455	saccharomyc	972	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
900	5	1.2	88	1	YRDB_ECOLI	P38455	saccharomyc	973	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
901	5	1.2	89	1	YRDB_ECOLI	P38455	saccharomyc	974	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
902	5	1.2	89	1	YRDB_ECOLI	P38455	saccharomyc	975	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
903	5	1.2	90	1	YRDB_ECOLI	P38455	saccharomyc	976	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
904	5	1.2	90	1	YRDB_ECOLI	P38455	saccharomyc	977	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
905	5	1.2	90	1	YRDB_ECOLI	P38455	saccharomyc	978	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
906	5	1.2	91	1	YRDB_ECOLI	P38455	saccharomyc	979	5	1.2	101	1	RS14_CHLNU	O961d4	chlamydia m
907	5	1.2	92	1	YRDB_ECOLI	P38455	saccharomyc	980	5	1.2	103	1	RS14_CHLNU	O961d4	chlamydia m
908	5	1.2	92	1	YRDB_ECOLI	P38455	saccharomyc	981	5	1.2	103	1	RS14_CHLNU	O961d4	chlamydia m
909	5	1.2	92	1	YRDB_ECOLI	P38455	saccharomyc	982	5	1.2	103	1	RS14_CHLNU	O961d4	chlamydia m

983 5 1.2 103 1 YTH4\_RHOER  
 984 5 1.2 104 1 C67B\_LOCMI  
 985 5 1.2 104 1 GRO\_CAVPO  
 986 5 1.2 104 1 HSP3\_RAT  
 987 5 1.2 104 1 PHE2\_RHOS2  
 988 5 1.2 104 1 S115\_HUMAN  
 989 5 1.2 105 1 AFI\_BOVIN  
 990 5 1.2 105 1 CORT\_HUMAN  
 991 5 1.2 105 1 FER\_SACER  
 992 5 1.2 105 1 PSBX\_GOSHI  
 993 5 1.2 105 1 THIO\_ALIAC  
 994 5 1.2 106 1 R33A\_YEAST  
 995 5 1.2 106 1 R33B\_YEAST  
 996 5 1.2 106 1 RLII\_HAEDU  
 997 5 1.2 106 1 TZAG\_ORYSA  
 998 5 1.2 106 1 Y04K\_BPT4  
 999 5 1.2 106 1 Y115\_ADE07  
 1000 5 1.2 107 1 GUNN\_CAVPO

## ALIGNMENTS

## RESULT 1

ITUA\_ECOLI STANDARD; PRT; 732 AA.  
 AC P14542;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ferric aerobactin receptor precursor (Cloacin receptor).  
 GN ITUA.  
 OS Escherichia coli.  
 OG Plasmid IncFI ColV3-K30.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=F344;  
 RA Krone W.J.A., Stegehuis F., Koningsstein G., van Doorn C.,  
 RA Roosendaal B., de Graaf F.K., Oudega B.;  
 RT "Characterization of the pColV-K30 encoded cloacin DF13/aerobactin  
 RT outer membrane receptor protein of Escherichia coli; isolation and  
 RT purification of the protein and analysis of its nucleotide sequence  
 RT and primary structure.";  
 RL FEMS Microbiol. Lett. 26:153-161(1985).  
 RN [2]  
 RP REVISIONS.  
 RA Oudega B.;  
 RL Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: RECEPTOR FOR CLOACIN DF13/AEROBACTIN.  
 CC -!- SUBCELLULAR LOCATION: Outer membrane.  
 CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.

-----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 -----

EMBL; X05874; CAA29297.1; -;  
 EMBL; X05874; CAA29298.1; ALT\_INIT.  
 PIR; S01042; S01042; TonB\_boxC.  
 InterPro; IPR000531; TonB\_boxC.  
 Pfam; PF00593; TonB\_boxC; 1.  
 PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; 1.  
 PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
 KW Outer membrane; Iron transport; TonB box; Signal; Plasmid;  
 Receptor.  
 FT SIGNAL 1 25

FT CHAIN 26 732 FERRIC AEROBACTIN RECEPTOR.  
 FT SITE 31 38 TONB\_BOX.  
 FT SITE 715 732 TONB\_C-TERMINAL\_BOX.  
 SQ SEQUENCE 732 AA; 81014 MW; 0C23879C0B27AE2B CRC64;

Query Match 2.5%; Score 10; DB 1; Length 732;

Best Local Similarity 100.0%; Pred. No. 0.052; Indels 0; Gaps 0;  
 Matches 10; Conservative 0; Mismatches 0;

Qy 154 TGGGLINIVTK 163

Db 144 TGGGLINIVTK 153

## RESULT 2

FXD3\_MOUSE STANDARD; PRT; 465 AA.  
 AC Q61050;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Forkhead box protein D3 (HNF3/FH transcription factor genesis)  
 DE (Hepatocyte nuclear factor 3 forkhead homolog 2) (HFN-2).  
 DE FOXD3 OR HFN2.  
 GN FOXD3 OR HFN2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hromas R.A., Costa R.H., Xu D., Sutton J.L.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Labosky P.A., Kaestner K.H.;  
 RT "The winged helix transcription factor Hfh2 is expressed in neural  
 RT crest and spinal cord during mouse development.";  
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 or send an email to license@isb-sib.ch).  
 -----

EMBL; U41047; AAA87569.1; -;  
 EMBL; AF067421; AAC28352.1; -;  
 HSSP; Q63245; 2HPH.  
 TRANSFAC; T04166; -;  
 MGD; MGI:1347473; Foxd3.  
 InterPro; IPR001766; TF\_Fork\_head.  
 Pfam; PF00250; Fork\_head; 1.  
 PRINTS; PR00053; FORKHEAD.  
 ProDom; PD000425; TF\_Fork\_head; 1.  
 SMART; SM00339; FH; 1.  
 PROSITE; PS00657; FORK\_HEAD\_1; 1.  
 PROSITE; PS00658; FORK\_HEAD\_2; 1.  
 PROSITE; PS00039; FORK\_HEAD\_3; 1.  
 KW DNA-binding; Nuclear protein; Transcription regulation.  
 FT DOMAIN 106 115 POLY-GLY.  
 FT DNA\_BIND 131 225 FORK-HEAD.  
 FT DOMAIN 252 257 POLY-ALA.  
 FT DOMAIN 265 270 POLY-ALA.  
 FT DOMAIN 275 281 POLY-ALA.  
 FT DOMAIN 380 399 POLY-GLY.  
 FT DOMAIN 447 457 POLY-ALA.  
 SQ SEQUENCE 465 AA; 47092 MW; 6F8B5B3D8C7564D CRC64;

Query Match 2.0%; Score 8; DB 1; Length 465;  
 Best Local Similarity 100.0%; Pred. No. 4.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 GSGATGGL 157  
 DB 113 GSGATGGL 120

## RESULT 3

DAG1\_BOVIN STANDARD; PRT; 895 AA.

AC 018738;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Dystroglycan precursor (Dystrophin-associated glycoprotein 1)  
 DE [Contains: Alpha-dystroglycan (Alpha-DG); Beta-dystroglycan (Beta-DG)]  
 GN DAG1.

OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 NC NCB1\_Taxid=9913;

(1)  
 SEQUENCE FROM N.A.

RA Shintzu H.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.

CHARACTERIZATION.  
 MEDLINE=99175209; PubMed=10075729;

RA Saito F., Masaki T., Kamakura K., Anderson L.V.B., Fujita S.,  
 RA Fukuta-Ohi H., Sunada Y., Shimizu T., Matsumura K.;  
 RT "Characterization of the transmembrane molecular architecture of the  
 RT dystroglycan complex in schwann cells.";  
 RL J. Biol. Chem. 274:8240-8246(1999).

-1- FUNCTION: FORMS PART OF THE DYSTROPHIN-ASSOCIATED PROTEIN COMPLEX  
 (DAPC) WHICH MAY LINK THE CYTOSKELETON TO THE EXTRACELLULAR  
 MATRIX. ALPHA-DYSTROGLYCAN FUNCTIONS AS A LAMININ RECEPTOR. BINDS  
 TO SEVERAL TYPES OF ARENAVIRUSES. IS A TARGET FOR THE ENTRY OF  
 MYCOBACTERIUM LEPRAE INTO PERIPHERAL NERVE SCHWANN CELLS.  
 CC -1- SUBCELLULAR LOCATION: ALPHA-DYSTROGLYCAN IS A EXTRACELLULAR  
 CC PROTEIN WHILE BETA-DYSTROGLYCAN IS A TYPE-1 MEMBRANE PROTEIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL, AB009079; BAA23650.1; -;  
 KW Signal; Glycoprotein; Transmembrane; Cytoskeleton.

FT SIGNAL 1 29 POTENTIAL.  
 FT CHAIN 30 653 ALPHA-DYSTROGLYCAN.  
 FT CHAIN 654 895 BETA-DYSTROGLYCAN.  
 FT DOMAIN 654 749 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 750 775 POTENTIAL.  
 FT DOMAIN 776 895 CYTOPLASMIC (POTENTIAL).  
 FT DISULFID 182 763 BY SIMILARITY.  
 FT DISULFID 669 713 POTENTIAL.  
 FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 641 641 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 649 649 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 661 661 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 895 AA; 97321 MW; 400213A299630D11 CRC64;

Query Match 2.0%; Score 8; DB 1; Length 895;  
 Best Local Similarity 100.0%; Pred. No. 7.6;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 LLSAVATQ 21  
 DB 21 LLSAVATQ 28

RESULT 4  
 NXS2\_LATCO STANDARD; PRT; 62 AA.

AC P10457;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Short neurotoxin II.  
 OS Laticauda colubrina (Yellow-lipped sea krait).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Lepidodonta; Squamata; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Lepidodonta; Squamata; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Elapidae; Laticaudinae; Laticauda.  
 NC NCB1\_Taxid=8628;  
 GN [1]  
 RP SEQUENCE.

RA STRAIN=Japanese, and Philippines; TISSUE=Venom;  
 RA Tamaya N., Sato A., Kim H.S., Teruchi T., Takasaki C., Ishikawa Y.,  
 RA Guinea M.L., McCoy M., Heatwole H., Cogger H.G.;  
 RT "Neurotoxins of sea snakes genus Laticauda.";  
 RT Toxicon 21 Suppl. 3:445-447(1983).

-1- FUNCTION: PRODUCES PERIPHERAL PARALYSIS BY BLOCKING NEUROMUSCULAR  
 TRANSMISSION AT THE POSTSYNAPTIC SITE. BINDS TO THE NICOTINIC  
 ACETYLCHOLINE RECEPTOR.

DR PIR; B25866; B25866.

DR HSSP; P01435; IQKE.  
 DR InterPro; IPR003571; Snake\_toxin.

DR Pfam; PF000087; toxin; 1.  
 DR ProDom; PD000206; Snake\_toxin; 1.

DR PROSITE; PS00272; SNAKE\_TOXIN; 1.  
 KW Venom; Neurotoxin; Postsynaptic neurotoxin; Multigene family.

FT DISULFID 3 24 BY SIMILARITY.  
 FT DISULFID 17 41 BY SIMILARITY.

FT DISULFID 43 54 BY SIMILARITY.  
 FT DISULFID 55 60 BY SIMILARITY.

SQ SEQUENCE 62 AA; 7037 MW; BCFP7BED2D71DDE CRC64;

Query Match 1.8%; Score 7; DB 1; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 7.2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 301 QPKTKS 307  
 DB 10 QPKTKS 16

RESULT 5  
 NXS2\_LATCO STANDARD; PRT; 62 AA.

AC P10455;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Short neurotoxin C.  
 OS Laticauda colubrina (Yellow-lipped sea krait).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Lepidodonta; Squamata; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Elapidae; Laticaudinae; Laticauda.  
 NC NCB1\_Taxid=8628;  
 GN [1]  
 RP SEQUENCE.

RA STRAIN=Salomon Island, and Fiji; TISSUE=Venom;  
 RA Tamaya N., Sato A., Kim H.S., Teruchi T., Takasaki C., Ishikawa Y.,  
 RA Guinea M.L., McCoy M., Heatwole H., Cogger H.G.;  
 RT "Neurotoxins of sea snakes genus Laticauda.";  
 RT Toxicon 21 Suppl. 3:445-447(1983).

-1- FUNCTION: PRODUCES PERIPHERAL PARALYSIS BY BLOCKING NEUROMUSCULAR  
 TRANSMISSION AT THE POSTSYNAPTIC SITE. BINDS TO THE NICOTINIC  
 ACETYLCHOLINE RECEPTOR.

CC

```
DR PIR; C25866; C25866.
DR HSP; P01435; IQKE.
DR InterPro; IPR003571; Snake_toxin.
DR Pfam; PF00087; toxin; 1.
DR ProDom; PD000206; Snake_toxin; 1.
DR PROSITE; PS00272; SNAKE_TOXIN; 1.
KW Venom; Neurotoxin; Postsynaptic neurotoxin; Multigene family.
FT DISULFID 3 24 BY SIMILARITY.
FT DISULFID 17 41 BY SIMILARITY.
FT DISULFID 43 54 BY SIMILARITY.
FT DISULFID 55 60 BY SIMILARITY.
SQ SEQUENCE 62 AA; 7064 MW; BCA94B3ED2D71D4E CRC64;

Query Match 1.8%; Score 7; DB 1; Length 62;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 301 QPKTTKS 307
DB 10 QPKTTKS 16

RESULT 6
ID _NXSD LATCO STANDARD; PRT; 62 AA.
AC P10456;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Short neurotoxin D.
OS Laticauda colubrina (Yellow-lipped sea krait).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodactylia; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Laticaudinae; Laticauda.
OX NCBI_TaxID=8628;
RN [1]
RP SEQUENCE.
RC STRAIN=New Caledonia; TISSUE=Venom;
RA Tamiya N., Sato A., Kim H.S., Teruuchi T., Takasaki C., Ishikawa Y.,
RA Guinea M.L., McCoy M., Heatwole H., Cogger H.G.;
RT "Neurotoxins of sea snakes genus Laticauda.";
RL Toxinon 21 Suppl. 3:445-447(1983)
CC -!- FUNCTION: PRODUCES PERIPHERAL PARALYSIS BY BLOCKING NEUROMUSCULAR
CC TRANSMISSION AT THE POSTSYNAPTIC SITE. BINDS TO THE NICOTINIC
CC ACETYLCHOLINE RECEPTOR.
DR PIR; D25866; D25866.
DR HSP; P01435; IQKE.
DR InterPro; IPR003571; Snake_toxin.
DR Pfam; PF00087; toxin; 1.
DR ProDom; PD000206; Snake_toxin; 1.
DR PROSITE; PS00272; SNAKE_TOXIN; 1.
KW Venom; Neurotoxin; Postsynaptic neurotoxin; Multigene family.
FT DISULFID 3 24 BY SIMILARITY.
FT DISULFID 17 41 BY SIMILARITY.
FT DISULFID 43 54 BY SIMILARITY.
FT DISULFID 55 60 BY SIMILARITY.
SQ SEQUENCE 62 AA; 7048 MW; BCA95421D2C8024E CRC64;

Query Match 1.8%; Score 7; DB 1; Length 62;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 301 QPKTTKS 307
DB 10 QPKTTKS 16

RESULT 7
GRPE CAUCR
ID _GRPE CAUCR STANDARD; PRT; 198 AA.
AC P48195;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
```

```
DT GRPE protein (HSP-70 cofactor).
DE GRPE OR CC0154.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CB15N / NA1000;
RC MEDLINE=96178944; PubMed=8606155;
RA Roberts R.C., Toochinda C., Avedissian M., Baldini R.L.,
RA Gomes S.L., Shapiro L.;
RT "Identification of a Caulobacter crescentus operon encoding hrcA,
RT involved in negatively regulating heat-inducible transcription, and
RT the chaperone gene grpB.";
RL J. Bacteriol. 178:1829-1841(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RC MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldlyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Skinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Fraser C.M.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -!- FUNCTION: STIMULATES, JOINTLY WITH DNAJ, THE ATPASE ACTIVITY OF
CC DNAK. HELPS TO RELEASE ADP FROM DNAK THUS ALLOWING DNAK TO RECYCLE
CC MORE EFFICIENTLY (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE GRPE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U33324; AAB01516.1; -.
CC EMBL; AE005689; AAK22141.1; -.
CC HSSP; P09372; 1DKG.
CC TIGR; CC0154; -.
CC InterPro; IPR000740; GrpE.
CC Pfam; PF01025; GrpE; 1.
CC PRINTS; PR00773; GRPEPROTEIN.
CC PROSITE; PS01071; GRPE; 1.
KW Chaperone; Heat shock; Complete proteome.
SQ SEQUENCE 198 AA; 20884 MW; 7BBEB334BCAPE368 CRC64;

Query Match 1.8%; Score 7; DB 1; Length 198;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 TGGAFDA 221
DB 190 TGGAFDA 196

RESULT 8
SODE HUMAN
ID _SODE HUMAN STANDARD; PRT; 240 AA.
AC P08234;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Extracellular superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1)
DE (EC-SOD).
DE GN SOD3.
```



OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RA Hjalmarsson K., Marklund S.L., Engstrom A., Edlund T.;  
 RT "Isolation and sequence of complementary DNA encoding human  
 extracellular superoxide dismutase.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:6340-6344(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Blood;  
 RX MEDLINE=95048365; PubMed=7959763;  
 RA Folz R.J., Crapo J.D.;  
 RT "Extracellular superoxide dismutase (SOD): tissue-specific  
 expression, genomic characterization, and computer-assisted sequence  
 analysis of the human EC SOD gene.";  
 RL Genomics 22:162-171(1994).  
 CC -1- FUNCTION: Destroys radicals which are normally produced within the  
 cells and which are toxic to biological systems.  
 CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).  
 CC -1- COFACTOR: Copper and zinc.  
 CC -1- SUBUNIT: HOMOTETRAMER.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- TISSUE SPECIFICITY: MAJOR SOD ISOZYME IN EXTRACELLULAR  
 FLUIDS SUCH AS PLASMA, LYMPH AND SYNOVIAL FLUID.  
 CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: J02947; AAA66000.1; -;  
 DR EMBL: U10116; AAA62278.1; -;  
 DR PIR: A28301; DSHUEC.  
 DR HSSP: P00442; ICBU.  
 DR Genew: HGNC:1181; SOD3.  
 DR MIM: 185490; -;  
 DR InterPro: IP0001424; SOD\_CU\_ZN.  
 DR Pfam: PF00080; sodcu.1  
 DR PRINTS: PR00068; CUZNDISMUTASE.  
 DR PROSITE: PS000469; SOD\_CU\_ZN\_1; 1.  
 DR PROSITE: PS00087; SOD\_CU\_ZN\_2; 1.  
 DR PROSITE: PS00332; SOD\_CU\_ZN\_2; 1.  
 DR Oxidoreductase; Copper; Zinc; Glycoprotein; Signal.  
 CC KEGG: K01101; SOD3.  
 FT CHAIN 1 18  
 FT SIGM 1 18  
 FT CHAIN 19 240  
 FT METAL 114 114  
 FT METAL 116 116  
 FT METAL 131 131  
 FT METAL 139 139  
 FT METAL 142 142  
 FT METAL 145 145  
 FT METAL 181 181  
 FT METAL 181 181  
 FT DISULFID 125 207  
 FT CARBOHYD 107 107  
 SO SEQUENCE 240 AA; 25881 MW; 5137848FC456344 CRC64;  
 Query Match 1.88; Score 7; DB 1; Length 240;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 9  
 ID CODE RABIT STANDARD; PRT; 244 AA.  
 AC P41975;  
 DT 01-NOV-1995 (rel. 32, Created)  
 DT 01-OCT-1996 (rel. 34, Last sequence update)  
 DT 15-JUN-2002 (rel. 41, Last annotation update)  
 DE Extracellular superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1)  
 DE (EC-SOD).  
 GN SOD3.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=New Zealand white;  
 RA Laakkonen M.O., Altmaki S.J., Hiltunen T.P., Yla-Herttuala S.;  
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=New Zealand white; TISSUE=Heart;  
 RA Laakkonen M.O., Hiltunen M.O., Altmaki S., Janne J.,  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Laakkonen M.O., Altmaki S., Mannermaa S., Hiltunen M.O.,  
 RL "Cloning and characterization of rabbit extracellular superoxide  
 dismutase.";  
 RT Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 43-155 FROM N.A.  
 RC STRAIN=New Zealand white; TISSUE=Aorta;  
 RA Hiltunen T.P., Nikkari T., Yla-Herttuala S.;  
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Destroys radicals which are normally produced within the  
 cells and which are toxic to biological systems.  
 CC -1- CATALYTIC ACTIVITY: 2 superoxide + 2 H(+) = O(2) + H(2)O(2).  
 CC -1- COFACTOR: Copper and zinc (by similarity).  
 CC -1- SUBUNIT: HOMOTETRAMER (by similarity).  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- SIMILARITY: BELONGS TO THE CU-ZN SUPEROXIDE DISMUTASE FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: Z67878; CAA91785.1; -;  
 DR EMBL: Y13339; CAA73783.1; -;  
 DR EMBL: AJ007044; CAA07431.1; -;  
 DR EMBL: X78139; CAA55018.1; -;  
 DR HSSP: P00442; ICBU.  
 DR InterPro: IP001424; SOD\_CU\_ZN.  
 DR Pfam: PF00080; sodcu.1.  
 DR PRINTS: PR00068; CUZNDISMUTASE.  
 DR PROSITE: PS000469; SOD\_CU\_ZN\_1; 1.  
 DR PROSITE: PS00087; SOD\_CU\_ZN\_2; 1.  
 DR PROSITE: PS00332; SOD\_CU\_ZN\_2; 1.  
 DR Oxidoreductase; Copper; Zinc; Glycoprotein; Signal.  
 CC KEGG: K01101; SOD3.  
 FT CHAIN 1 18  
 FT SIGM 1 18  
 FT CHAIN 19 244  
 FT METAL 118 118  
 FT METAL 120 120  
 FT METAL 135 135  
 FT METAL 143 143  
 FT METAL 143 143  
 SO SEQUENCE 244 AA; 25881 MW; 5137848FC456344 CRC64;  
 Query Match 1.88; Score 7; DB 1; Length 240;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
FT METAL 146 146 ZINC (BY SIMILARITY).
FT METAL 149 149 ZINC (BY SIMILARITY).
FT METAL 185 185 COPPER (BY SIMILARITY).
FT DTULFD 129 211 BY SIMILARITY.
FT CARBOHYD 111 111 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CONFLICT 53 53 R -> W (IN REF. 4).
SQ SEQUENCE 244 AA; 25688 MW; 7C9B1C59F942F2C5 CRC64;

Query Match 1.8%; Score 7; DB 1; Length 244;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 358 PSATLDA 364
Db 71 PSATLDA 77

RESULT 10
1431 SCHMA STANDARD; PRT; 252 AA.
AC Q26540;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE 14-3-3 protein homolog 1.
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea; Strigeidida;
OC Schistosomatidae; Schistosomatidae; Schistosoma.
OX NCBI_TaxID=6183;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Puerto Rican; PubMed=8577340;
RX MEDLINE=96123403; PubMed=8577340;
RA Schechterman D., Ram D., Tarrab-Hazdai R., Arnon R., Schechter I.;
RT "Stage-specific expression of the mRNA encoding a 14-3-3 protein
during the life cycle of Schistosoma mansoni.";
RL Mol. Biochem. Parasitol. 73:275-278(1995).
CC -!- SIMILARITY: BELONGS TO THE 14-3-3 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U24281; AAC46983.1; -.
DR HSSP; P29312; 1A38.
DR InterPro; IPR000308; 14-3-3.
DR Pfam; PF00244; 14-3-3; 1.
DR PRINTS; PR00305; 1433ZETA.
DR ProDom; PD000600; 14-3-3; 1.
DR SMART; SM00101; 1433_3; 1.
DR PROSITE; PS00796; 1433_1; 1.
DR PROSITE; PS00797; 1433_2; 1.
KW Multigene family.
FT SIMILAR 133 149 TO THE C-TERMINUS OF ANNEXINS.
SQ SEQUENCE 252 AA; 28371 MW; 39103EB5B71A2C08 CRC64;

Query Match 1.8%; Score 7; DB 1; Length 252;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 362 LDAYTKA 368
Db 157 LDAYTKA 163

RESULT 11
PANE_VIBCH STANDARD; PRT; 264 AA.
ID PANE_VIBCH
AC Q9KUD0;

DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 3-methyl-2-oxobutanate hydroxymethyltransferase (EC 2.1.2.11)
DE (ketopantoate hydroxymethyltransferase).
DE PANB OR VC0592
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
Emmolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae.";
RL Nature 406:477-483(2000).
CC -!- CATALYTIC ACTIVITY: 5.10-methylenetetrahydrofolate + 3-methyl-2-
oxobutanate = tetrahydrofolate + 2-dehydropanoate.
CC -!- PATHWAY: Pantothenate biosynthesis; first branch; first step.
CC -!- SIMILARITY: BELONGS TO THE PANB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE004144; AAF93759.1; -.
DR TIGR; VC0592; -.
DR InterPro; IPR003700; Pantoate transf.
DR Pfam; PF02548; Pantoate transf; 1.
DR TIGRPFAM; TIGR00222; panB; 1.
KW Pantothenate biosynthesis; Transferase; Methyltransferase;
Complete proteome.
SQ SEQUENCE 264 AA; 28659 MW; 68A61A0D979351B3 CRC64;

Query Match 1.8%; Score 7; DB 1; Length 264;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 50 LANRITQ 56
Db 187 LANRITQ 193

RESULT 12
NTCP_RAT STANDARD; PRT; 362 AA.
AC P26435;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Sodium/bile acid cotransporter (Na(+)/bile acid cotransporter)
DE (Na(+)/taurocholate transport protein) (Sodium/taurocholate
cotransporting polypeptide).
GN SLC10A1 OR NTCP
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=92073340; PubMed=1961729;
```

RA Hagenbuch B., Stieger B., Foguet M., Luebbert H., Meier P.J.;  
 RT "Functional expression cloning and characterization of the hepatocyte  
 Na+/bile acid cotransport system";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:10629-10633(1991).  
 CC -1- FUNCTION: THE HEPATIC SODIUM/BILE ACID UPTAKE SYSTEM EXHIBITS  
 CC BROAD SUBSTRATE SPECIFICITY & TRANSPORTS VARIOUS NONBILE ACID  
 CC ORGANIC COMPOUNDS AS WELL. IT IS STRICTLY DEPENDENT ON THE  
 CC EXTRACELLULAR PRESENCE OF SODIUM.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- TISSUE SPECIFICITY: LIVER AND KIDNEY.  
 CC -1- SIMILARITY: BELONGS TO THE SODIUM-BILE ACID SYMPORTER FAMILY  
 CC (SBR).  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; W77479; AAA2112.1; -  
 CC PIR; A41601; A41601.  
 DR InterPro; IPR004710; Baas.  
 DR InterPro; IPR002657; BileAc/Na\_smptr.  
 DR Pfam; PF01758; SBR; 1.  
 DR TIGRPFAM; TIGR00841; baas; 1.  
 KW Transmembrane; Transport; Symport; Sodium transport; Glycoprotein.  
 FT TRANSMEM 24 45  
 FT TRANSMEM 60 80 POTENTIAL.  
 FT TRANSMEM 82 98 POTENTIAL.  
 FT TRANSMEM 158 178 POTENTIAL.  
 FT TRANSMEM 190 211 POTENTIAL.  
 FT TRANSMEM 228 244 POTENTIAL.  
 FT TRANSMEM 285 306 POTENTIAL.  
 FT CARBOHYD 5 306 POTENTIAL.  
 FT CARBOHYD 11 11 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 117 117 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SO SEQUENCE 362 AA; 39295 MW; F0ABB76076A57550 CRC64;  
 -----  
 Query Match 1.8%; Score 7; DB 1; Length 362;  
 Best Local Similarity 100.0%; Pred. No. 36;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 14 LISVAAT 20  
 DB 197 LISVAAT 203  
 -----  
 RESULT 13  
 ID GP17\_HUMAN STANDARD; PRT; 367 AA.  
 AC Q1304; Q9UDZ6; Q9UE21;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DE Probable P2Y purinoceptor GPR17 (P2Y-like receptor) (R12).  
 GN GPR17.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX MEDLINE=96145150; PubMed=8558062;  
 RA Raport C.J., Schweickart V.L., Chantry D., Eddy R.L., Jr., Shows T.B.,  
 RA Godolka R., Gray P.W.;  
 RT "New members of the chemokine receptor gene family";  
 RT J. Leukoc. Biol. 59:18-23(1996).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).

RC TISSUE=Hippocampus;  
 RX MEDLINE=98181695; PubMed=9523551;  
 RA Blasius R.H., Weber R.G., Lichter P., Ogilvie A.;  
 RT "A novel orphan G protein-coupled receptor primarily expressed in the  
 RT brain is localized on human chromosomal band 2q21.";  
 RL J. Neurochem. 70:1357-1365(1998).  
 CC -1- FUNCTION: Putative receptor for purines coupled to G-proteins (By  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are  
 CC produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN BRAIN.  
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 CC EMBL; U33447; AAB16746.1; -  
 CC EMBL; Y12546; CAA73144.1; -  
 DR EMBL; Z94154; CAB08107.1; -  
 DR EMBL; Z94155; CAB08108.1; -  
 DR HSSP; P34996; 1DD0.  
 DR Genew; HGNC:4471; GPR17.  
 DR MIM; 603071; -  
 DR InterPro; IPR000276; GPCR\_Rhodpsn.  
 DR Pfam; PF00001; 7tm\_1; 1.  
 DR PRINTS; PR00237; GPCRHHODPSN.  
 DR PROSITE; PS00237; G-PROTEIN\_REC\_P1\_1; 1.  
 DR PROSITE; PS00262; G-PROTEIN\_REC\_P1\_2; 1.  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;  
 KW Alternative splicing.  
 FT DOMAIN 1 64  
 FT TRANSMEM 65 85  
 FT DOMAIN 86 92  
 FT TRANSMEM 93 113  
 FT DOMAIN 114 133  
 FT TRANSMEM 134 154  
 FT DOMAIN 155 175  
 FT TRANSMEM 176 196  
 FT DOMAIN 197 223  
 FT TRANSMEM 224 244  
 FT DOMAIN 245 260  
 FT TRANSMEM 261 281  
 FT DOMAIN 282 308  
 FT TRANSMEM 309 329  
 FT DOMAIN 330 367  
 FT DISULFID 132 209  
 FT CARBOHYD 42 42  
 FT CARBOHYD 204 204 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 282 282 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPPLIC 1 28 MISSING (IN ISOFORM 2).  
 SO SEQUENCE 367 AA; 40989 MW; 132FBB97BB8360C CRC64;  
 -----  
 Query Match 1.8%; Score 7; DB 1; Length 367;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 49 ALANRIT 55  
 DB 304 ALANRIT 310  
 -----  
 RESULT 14  
 ID POLG\_MDMV STANDARD; PRT; 380 AA.  
 AC P32652;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Genome polyprotein [Contains: Nuclear inclusion protein B (NI-B) (NIB)  
DE (RNA-directed RNA polymerase) (EC 2.7.7.48); Coat protein (CP)]  
DE (Fragment).  
OS Maize dwarf mosaic virus (MDMV).  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;  
OC Potyvirus.  
OX NCBI\_TaxID=12203;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B;  
RX MEDLINE=91132116; PubMed=1993866;  
RA Frenkel M.J., Jilka J.M., McKern N.M., Strike P.M., Clark J.M. Jr.,  
RA Shukla D.D., Ward C.W.;  
RA "Unexpected sequence diversity in the amino-terminal ends of the coat  
RT proteins of strains of sugarcane mosaic virus";  
RL J. Gen. Virol. 72:237-242(1991).  
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
CC {RNA} (N).  
CC -1- PTM: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE  
CC POLYPEPTIDE WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC  
CC PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT  
CC INDIVIDUAL PROTEINS.  
CC -1- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPEPTIDE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; D00949; BAA00797.1; -;  
DR PIR; PH0208; GNUSMB.  
DR InterPro; IPR001592; Poty\_coat.  
DR Pfam; PF00767; Poty\_coat; 1.  
DR Transferase; RNA-directed RNA polymerase; Coat protein; Polyprotein.  
FT NON\_TER 1 1  
FT CHAIN 53 380 NUCLEAR INCLUSION PROTEIN B.  
FT CHAIN 53 380 COAT PROTEIN.  
SQ SEQUENCE 380 AA; 41119 MW; BBAFP2AA45E1CB82 CRC64;  
Query Match 1.8%; Score 7; DB 1; Length 380;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 150 GSGATGG 156  
Db 111 GSGATGG 117  
RESULT 15  
AMPH\_ECOLI STANDARD; PRT; 385 AA.  
AC P46127; P75701;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Penicillin-binding protein amph.  
GN AMPH OR B0376 OR Z0472 OR ECS0426.  
OS Escherichia coli, and  
OS Escherichia coli O157:H7.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia  
OX NCBI\_TaxID=562, 83334;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

MAU B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12";  
RL Science 277:1453-1474 (1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Duncan M., Allen B., Araujo R., Aparicio A.M., Chung B., Davis K.,  
RA Pederspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H.,  
RA Lin D., Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;  
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
RX MEDLINE=21074955; PubMed=11206551;  
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RL "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";  
RN Nature 409:529-533(2001).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
RL "Complete genome sequence of enterohaemorrhagic Escherichia coli  
RT O157:H7 and genomic comparison with a laboratory strain K-12";  
RN DNA Res. 8:11-22(2001).  
RN [5]  
RP SEQUENCE OF 1-78 FROM N.A.  
RC STRAIN=K12;  
RX Moreno F.;  
RL Submitted (JUL-1990) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP IDENTIFICATION.  
RX MEDLINE=96032851; PubMed=7567469;  
RA Borodovsky M., McIninch J., Koonin E.V., Rudd K.E., Medigue C.,  
RA Danchin A.;  
RT "Detection of new genes in a bacterial genome using Markov models for  
RL three gene classes";  
RN Nucleic Acids Res. 23:3554-3562(1995).  
RN [7]  
RP CHARACTERIZATION.  
RX MEDLINE=97464439; PubMed=9324260;  
RA Henderson T.A., Young K.D., Denome S.A., Elf P.K.;  
RT "AmpC and AmpH, proteins related to the class C beta-lactamases, bind  
RL penicillin and contribute to the normal morphology of Escherichia  
RN coli";  
RP J. Bacteriol. 179:6112-6121(1997).  
CC -1- FUNCTION: BINDS PENICILLIN AND CONTRIBUTES TO THE NORMAL  
CC MORPHOLOGY OF ESCHERICHIA COLI.  
CC -1- SIMILARITY: TO CLASS-C BETA-LACTAMASES.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE000144; AAC73479.1; -;  
DR EMBL; U73857; AAB18099.1; ALT\_INIT.  
DR EMBL; AE005216; AAG54722.1; -;  
DR EMBL; AP002551; BAB33849.1; -;  
DR EMBL; X54153; -; NOT\_ANNOTATED\_CDS.  
DR MEROPS; S12.UNW; -;  
DR EcoGene; EGI2867; amph.  
KW Complete proteome.

SO SEQUENCE 385 AA; 41849 MW; 2D249CBA78022947 CRC64;

Query Match 1.8%; Score 7; DB 1; Length 385;

Best Local Similarity 100.0%; Pred. No. 38;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 YGSGATG 155

DB 44 YGSGATG 50

RESULT 16

MUTL\_AQUPY STANDARD; PRT; 426 AA.

ID MUTL\_AQUPY

AC P70754;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE DNA mismatch repair protein mutL.

GN MUTL.

OS Aquifex pyrophilus.

OC Bacteri; Aquificae; Aquificae (class); Aquificales; Aquificaceae;

OC Aquifex

NCBI\_TaxID=2714;

RN (1)

SEQUENCE FROM N.A.

RA Wetmur J.G., Rosenfeld A., Wong D.M.;

RT "Expression and characterization of MutL proteins from thermophilic

subbacteria";

RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN THE REPAIR OF MISMATCHES IN

CC DNA. IT IS REQUIRED FOR DAM-DEPENDENT METHYL-DIRECTED DNA MISMATCH

CC REPAIR. MAY ACT AS A "MOLECULAR MATCHMAKER", A PROTEIN THAT

CC PROMOTES THE FORMATION OF A STABLE COMPLEX BETWEEN TWO OR MORE

CC DNA-BINDING PROTEINS IN AN ATP-DEPENDENT MANNER WITHOUT ITSELF

CC BEING PART OF A FINAL EFFECTOR COMPLEX (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTL/HEXB FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; U71052; AAB09595.1; -

DR HSSP; P23367; 1BKX.

DR InterPro; IPR003594; ATPbind\_ATPase.

DR InterPro; IPR002099; DNA\_mis\_repair.

DR InterPro; IPR004359; HIS\_KIN\_slg.

DR Pfam; PF01119; DNA\_mis\_repair; 1.

DR Pfam; PF02518; HATase\_c; 1.

DR TIGRFAMs; TIGR00585; mutL; 1.

DR PROSITE; PS00058; DNA\_MISMATCH\_REPAIR\_1; 1.

KW DNA repair.

SO SEQUENCE 426 AA; 49791 MW; D81943BE45E7D4 CRC64;

Query Match 1.8%; Score 7; DB 1; Length 426;

Best Local Similarity 100.0%; Pred. No. 42;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 141 VLSGATS 147

DB 72 VLSGATS 78

RESULT 17

OADB\_SALTY STANDARD; PRT; 433 AA.

ID OADB\_SALTY

AC Q829K6;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DB Oxaloacetate decarboxylase beta chain (EC 4.1.1.3).

OADB OR STY3531.

OS Salmonella typhi.

OC Bacteri; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Salmonella.

NCBI\_TaxID=601;

RN (1)

SEQUENCE FROM N.A.

RA STRAIN=CT18;

RC MEDLINE=21534947; PubMed=11677608;

RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,

RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,

RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,

RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,

RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jorgels K.,

RA Krogh A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,

RA Quail M., Rutherford K., Simmonds M., Skellon J., Stevens K.,

RA Whitehead S., Barrall B.G.;

RT "Complete genome sequence of a multiple drug resistant Salmonella

RT enterica serovar Typhi CT18.";

RL Nature 413:848-852(2001).

CC -1- FUNCTION: LYSASE AND SODIUM TRANSPORTER.

CC -1- CATALYTIC ACTIVITY: Oxaloacetate = pyruvate + CO(2).

CC -1- COFACTOR: REQUIRES A SODIUM ION.

CC -1- SUBUNIT: HETEROTRIMER OF AN ALPHA, A BETA AND A GAMMA SUBUNIT.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

CC -1- SIMILARITY: BELONGS TO THE GCDB / MMDB / OADB FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AL627265; CAD01209.1; -

DR TIGRFAMs; TIGR01109; Na\_pump\_decarb; 1.

KW Decarboxylase; Transmembrane; Lyase; Sodium transport;

KW Complete proteome.

FT TRANSMEM 13 35 POTENTIAL.

FT TRANSMEM 42 64 POTENTIAL.

FT TRANSMEM 125 147 POTENTIAL.

FT TRANSMEM 160 182 POTENTIAL.

FT TRANSMEM 214 236 POTENTIAL.

FT TRANSMEM 266 288 POTENTIAL.

FT TRANSMEM 308 327 POTENTIAL.

FT TRANSMEM 339 361 POTENTIAL.

FT TRANSMEM 413 432 POTENTIAL.

SO SEQUENCE 433 AA; 44939 MW; 8B06DC6019B84357 CRC64;

Query Match 1.8%; Score 7; DB 1; Length 433;

Best Local Similarity 100.0%; Pred. No. 43;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 156 GLNIVT 162

DB 312 GLNIVT 318

RESULT 18

OADB\_SALTY STANDARD; PRT; 433 AA.

ID OADB\_SALTY

AC Q030I1;

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Oxaloacetate decarboxylase beta chain (EC 4.1.1.3).

GN OADB OR STM3351.

OS Salmonella typhimurium.

OC Bacteri; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Salmonella.

```

NCBI_TaxID=602;
[1]
SEQUENCE FROM N.A.
STRAIN=LT2;
MEDLINE=93054591; PubMed=1331067;
RA Woehle G., Wifling K., Dmroch P.;
RT "Sequence of the sodium ion pump oxaloacetate decarboxylase from
RL Salmonella typhimurium.";
J. Biol. Chem. 267:22798-22803(1992).
[2]
SEQUENCE FROM N.A.
STRAIN=LT2 / SGSC1412 / ATCC 700720;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney S., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RL LT2.";
Nature 413:852-856(2001).
CC -1- FUNCTION: LYASE AND SODIUM TRANSPORTER.
CC -1- CATALYTIC ACTIVITY: Oxaloacetate = pyruvate + CO(2).
CC -1- COFACTOR: REQUIRES A SODIUM ION.
CC -1- SUBUNIT: HETEROTRIMER OF AN ALPHA, A BETA AND A GAMMA SUBUNIT.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE GCDP / MMDB / OADB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M96434; AAA02974.1; -.
DR EMBL; AE008854; AAL22220.1; -.
DR PIR; C44465; C44465.
DR STYGene; SG10260; OADB.
DR TIGRfams; TIGR01109; Na_pump_decarbB; 1.
KW Decarboxylase; Transmembrane; Lyase; Sodium transport;
KW Complete proteome.
FT TRANSMEM 13 35 POTENTIAL.
FT TRANSMEM 42 64 POTENTIAL.
FT TRANSMEM 125 147 POTENTIAL.
FT TRANSMEM 160 182 POTENTIAL.
FT TRANSMEM 214 236 POTENTIAL.
FT TRANSMEM 266 288 POTENTIAL.
FT TRANSMEM 308 327 POTENTIAL.
FT TRANSMEM 339 361 POTENTIAL.
FT TRANSMEM 413 432 POTENTIAL.
SQ SEQUENCE 433 AA; 44923 MW; 1B29B1376B4632A1 CRC64;

Query Match 1.8%; Score 7; DB 1; Length 433;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 156 GLINIVT 162
DB 312 GLINIVT 318

RESULT 19
LEF4_NPVP
ID LEF4_NPVP STANDARD; PRT; 457 AA.
AC O10340;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Late expression factor 4.
GN LEF-4.
OS Orgyia pseudotsugata multicapsid polyhedrosis virus (OpMNPV).
```

```

Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
Nucleopolyhedrovirus.
NCBI_TaxID=164623;
[1]
SEQUENCE FROM N.A.
MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA Rohrmann G.F.;
RT "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear
RL polyhedrosis virus genome.";
Virology 229:381-399(1997).
CC -1- FUNCTION: REQUIRED FOR LATE AND VERY LATE GENE EXPRESSION (BY
CC SIMILARITY).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U75930; AAC59090.1; -.
KW Transcription regulation
SQ SEQUENCE 457 AA; 51151 MW; E21E06E50BAC8390 CRC64;

Query Match 1.0%; Score 7; DB 1; Length 457;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 31 LPTVELE 37
DB 394 LPTVELE 400

RESULT 20
VS11_REOVJ
ID VS11_REOVJ STANDARD; PRT; 462 AA.
AC P04507;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Sigma 1 protein precursor (Hemagglutinin) (Cell attachment protein)
DE (Minor outer capsid protein).
DE SI.
OS Reovirus (type 2 / strain D5/Jones).
OC Viruses; dsRNA viruses; Reoviridae; Orthoreovirus.
OX NCBI_TaxID=10885;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90163207; PubMed=2305549;
RA Duncan R., Horne D., Cashdollar L.W., Joklik W.K., Lee P.W.K.;
RT "Identification of conserved domains in the cell attachment proteins
RL of the three serotypes of reovirus.";
Virology 174:399-409(1990).
[2]
SEQUENCE FROM N.A.
RX MEDLINE=85113159; PubMed=3855545;
RA Cashdollar L.W., Chmelo R.A., Wiener J.R., Joklik W.K.;
RT "Sequences of the SI genes of the three serotypes of reovirus.";
Proc. Natl. Acad. Sci. U.S.A. 82:24-28(1985).
CC -1- SUBUNIT: Tetramer.
CC -1- SIMILARITY: 25% TO SEROTYPE 3 AND 48% TO SEROTYPE 1 SIGMA 1
CC PROTEINS.
CC
CC -1- CAUTION: SIGMA WAS RESEQUENCED (REF.1) AND SHOWS DIFFERENCES TO
CC THE SEQUENCE PUBLISHED PREVIOUSLY (REF.2).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```

CC or send an email to license@ibb-sib.ch).
-----
DR EMBL; M32861; AAA47268.1; -.
DR EMBL; M10261; AAA66879.1; ALT_SEQ.
DR PIR; A04123; HXXRH2.
DR InterPro; IPR002592; Rec_sigmal.
KW Pfam; PF01664; Rec_sigmal; 1.
Coat protein; Glycoprotein; Hemagglutinin; signal.
FT SIGNAL 1
FT CHAIN 1
FT CARBOHYD 237 237 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 462 AA; 50479 MW; 964A3E4F65ACE74 CRC64;

Query Match
Best Local Similarity 1.8%; Score 7; DB 1; Length 462;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 197 GVSSENGN 203
Db 69 GVSSENGN 75

RESULT 21
UCR2_EUGGR STANDARD; PRT; 474 AA.
AC P43265;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ubiquitinol-cytochrome C reductase complex core protein 2, mitochondrial
DE precursor (EC 1.10.2.2).
OS Euglena gracilis.
OC Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.
OX NCBI_TaxID=3039;
RN (1)
RP SEQUENCE FROM N.A., AND SEQUENCE OF 43-54.
RC STRAIN=SM-ZK;
RX MEDLINE=94245672; PubMed=8186644;
RA Chit J.-Y., Mukai K., Saeki K., Matubara H.;
RT "Molecular cloning and nucleotide sequences of cDNAs encoding
RT subunits I, II, and IX of Euglena gracilis mitochondrial complex
RT III."
RT J. Biochem. 115:98-107(1994).
CC -1- FUNCTION: THIS IS A COMPONENT OF THE UBIQUINOL-CYTOCHROME C
CC REDUCTASE COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH
CC IS PART OF THE MITOCHONDRIAL RESPIRATORY CHAIN. THE CORE PROTEIN
CC 2 IS REQUIRED FOR THE ASSEMBLY OF THE COMPLEX.
CC -1- CATALYTIC ACTIVITY: OX(2) + 2 ferriocytochrome c = O + 2
CC ferriocytochrome c.
CC -1- SUBCELLULAR LOCATION: Mitochondrial inner membrane; matrix side.
CC -1- CAUTION: DOES NOT SEEM TO HAVE A PROTEASE ACTIVITY AS IT LACK THE
CC ZINC-BINDING SITE.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ibt-sib.ch/announce/
CC or send an email to license@ibb-sib.ch).
CC -----
DR EMBL; D16672; BAA04080.1; -.
DR MEROPS; M16; UNB; -.
DR InterPro; IPR001431; Peptidase M16.
DR Pfam; PF00675; Peptidase M16; 1.
DR PROSITE; PS00143; INSULINASE; FALSE NEG.
KW Mitochondrion; Inner membrane; Electron transport; Respiratory chain;
KW Oxidoreductase; Transit peptide.
FT TRANSIT 1
FT MITOCHONDRION.

```

```

FT CHAIN 43 474 UBIQUINOL-CYTOCHROME C REDUCTASE COMPLEX
FT SEQUENCE 474 AA; 51073 MW; 4CDF5C27AF119175 CRC64;
SQ SEQUENCE

Query Match
Best Local Similarity 1.8%; Score 7; DB 1; Length 474;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 SINPNV 136
Db 24 SINPNV 30

RESULT 22
MUTL_CHLTR STANDARD; PRT; 576 AA.
AC 084579;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA mismatch repair protein mutL.
GN MUTL OR CTS75.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/Cx;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marache R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RT Science 282:754-759(1998).
CC -1- FUNCTION: THIS PROTEIN IS INVOLVED IN THE REPAIR OF MISMATCHES IN
CC DNA. IT IS REQUIRED FOR DAM-DEPENDENT METHYL-DIRECTED DNA MISMATCH
CC REPAIR. MAY ACT AS A "MOLECULAR MATCHMAKER", A PROTEIN THAT
CC PROMOTES THE FORMATION OF A STABLE COMPLEX BETWEEN TWO OR MORE
CC DNA-BINDING PROTEINS IN AN ATP-DEPENDENT MANNER WITHOUT ITSELF
CC BEING PART OF A FINAL EFFECTOR COMPLEX (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE DNA MISMATCH REPAIR MUTL/HEXB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ibt-sib.ch/announce/
CC or send an email to license@ibb-sib.ch).
CC -----
DR EMBL; AE001328; AAC68177.1; -.
DR HSSP; P23367; IBKN.
DR InterPro; IPR003594; ATPbind ATPase.
DR InterPro; IPR002099; DNA_mis_repair.
DR Pfam; PF01119; DNA_mis_repair; 1.
DR Pfam; PF02518; HATPase_C; 1.
DR TIGRfam; TIGR00585; mutL; 1.
DR PROSITE; PS00058; DNA_MISMATCH_REPAIR_1; 1.
KW DNA repair; Complete proteome.
SQ SEQUENCE 576 AA; 64138 MW; 8AAF6CC19783F84E CRC64;

Query Match
Best Local Similarity 1.8%; Score 7; DB 1; Length 576;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 LSSEGIG 188
Db 183 LSSEGIG 189

RESULT 23
IF2_MYCPN

```

```
ID AC IF2 MYCPN STANDARD; PRT; 617 AA.
AC P75590;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Translation initiation factor IF-2.
GN INF2 OR MN155 OR MP676.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreich R., Hilbert H., Piagens H., Pirkl E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- FUNCTION: IF-2, ONE OF THE ESSENTIAL COMPONENTS FOR THE INITIATION
CC OF PROTEIN SYNTHESIS IN VITRO, PROTECTS FORMYLMETHIONYL-TRNA FROM
CC SPONTANEOUS HYDROLYSIS AND PROMOTES ITS BINDING TO THE 30S
CC RIBOSOMAL SUBUNITS. IT IS ALSO INVOLVED IN THE HYDROLYSIS OF GTP
CC DURING THE FORMATION OF THE 70S RIBOSOMAL COMPLEX (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000062; AAB96324.1; -
DR InterPro; IPR004161; EFTU_D2.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR000178; IF2.
DR InterPro; IPR005225; Small GTP.
DR Pfam; PF00009; GTP_EFTU_1.
DR Pfam; PF03144; GTP_EFTU_D2; 2.
DR ProDom; PD186100; IF2; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR TIGRFAMs; TIGR00487; IF-2; 1.
DR PROSITE; PS01176; IF2; 1.
DR Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT DOMAIN 122 271 G-DOMAIN.
FT NP_BIND 128 135 GTP (BY SIMILARITY).
FT NP_BIND 175 179 GTP (BY SIMILARITY).
FT NP_BIND 229 232 GTP (BY SIMILARITY).
SQ SEQUENCE 617 AA; 67897 MW; B658142123D70133 CRC64;

Query Match 1.8%; Score 7; DB 1; Length 617;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Gaps 0;

QY 113 LNGVPLT 119
DB 66 LNGVPLT 72

RESULT 24
PTAA_ECOLI
ID PTAA_ECOLI STANDARD; PRT; 648 AA.
AC P09323;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE PTS system, N-acetylglucosamine-specific IIABC component (EIIABC-NAG)
DE (N-acetylglucosamine-permease IIABC component) (Phosphotransferase
DE enzyme II, ABC component) (EC 2.7.1.69) (EII-NAG).
```

```
GN OS NAGE OR PSTN OR B0679.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88212176; PubMed=3284790;
RA Rogers M.J., Ongi T., Plumbridge J., Soell D.;
RT "Nucleotide sequences of the Escherichia coli nagE and nagB genes:
RT the structural genes for the N-acetylglucosamine transport protein of
RT and the bacterial phosphoenolpyruvate: sugar phosphotransferase system
RT and for glucosamine-6-phosphate deaminase.";
RL Gene 62:197-207(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89050950; PubMed=3056518;
RA Peri K.G., Waygood E.B.;
RT "Sequence of cloned enzyme IIN-acetylglucosamine of the
RT phosphoenolpyruvate:N-acetylglucosamine phosphotransferase system of
RT Escherichia coli.";
RL Biochemistry 27:6054-6061(1988).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glaesner J.D., Rode C.K., Mayhew G.P.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HEP); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR.
CC -!- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC histidine + sugar phosphate.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC -!- SIMILARITY: CONTAINS 1 PTS EIIA DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PTS EIIB DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PTS EIIC DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; M19284; AAA24192.1; -
EMBL; AE000171; AAC73773.1; -
EMBL; D90706; BAA35322.1; -
EMBL; D90707; BAA35327.1; -
PIR; B29895; WOEC2N.
PIR; A28896; A28896.
```







```

RP SEQUENCE.
RC TISSUE=Mesocarp; PubMed=9022685;
RX MEDLINE=97175025; PubMed=9022685;
RA Yoshinari S., Yokota S., Sawmoto H., Koresawa S., Tamura M.,
RA Endo Y.;
RT "Purification, characterization and subcellular localization of a
RT type-1 ribosome-inactivating protein from the sarcocarp of Cucurbita
RT pepo.";
RL Eur. J. Biochem. 242:585-591(1996).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of the N-glycosidic bond at one
CC specific adenosine on the 28S rRNA.
CC -1- SIMILARITY: BELONGS TO THE RIBOSOME-INACTIVATING PROTEIN FAMILY.
CC TYPE 1 RIP SUBFAMILY.
CC InterPro: IPR001574; RIP.
DR Pfam: PF00161; RIP; 1.
DR PROSITE: PS00275; SHIGA_RICIN; PARTIAL.
KM Plant defense: Protein synthesis inhibitor; Hydrolase; Toxin.
FT NON TER 20 20
SQ SEQUENCE 20 AA; 2235 MW; 8E753D217FDEA6D CRC64;

Query Match 1.5%; Score 6; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 LSGATS 147
Db 6 LSGATS 11

RESULT 30
PEN3 ADECU STANDARD; PRT; 33 AA.
ID PEN3 ADECU
AC P35987;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Penton protein (Varion component III) (Penton base protein)
DE (Fragment).
DE PIII.
GN PIII.
OS Canine adenovirus type 1 (strain Utrecht).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=36364;
RN NCBI_TaxID=36364;
RP SEQUENCE FROM N.A.
RX MEDLINE=93212535; PubMed=8460501;
RA Cai F., Weber J.M.;
RT "Primary structure of the canine adenovirus PVI protein: functional
RT implications.";
RL Virology 193:986-988(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: W73811; AAA75345.1; -
DR PIR: A46116; A46116.
DR InterPro: IPR002605; Adeno_Penton_B.
DR Pfam: PF01686; Adeno_Penton_B; 1.
KM Late protein.
FT NON TER 1 1
SQ SEQUENCE 33 AA; 3633 MW; 06CAF4F39EBB6104A CRC64;

Query Match 1.5%; Score 6; DB 1; Length 33;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 87 IPELGV 92
Db 17 IPELGV 22

```

```

RESULT 31
RIP2 TRIKI STANDARD; PRT; 45 AA.
ID RIP2 TRIKI
AC P23029;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Ribosome-inactivating protein TAP-29 (rRNA N-glycosidase)
DE (EC 3.2.2.22) (Fragment).
DE Trichosanthes kirilowii (Mongolian snake-gourd).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Cucurbitales; Cucurbitaceae; Trichosanthes.
OX NCBI_TaxID=3677;
RN NCBI_TaxID=3677;
RP SEQUENCE.
RC TISSUE=Tuberous root;
RX MEDLINE=91319727; PubMed=1713684;
RA Lee-Huang S., Huang P.L., Kung H.-F., Li B.-Q., Huang P.L., Huang P.,
RA Huang H.I., Chen H.-C.;
RT "TAP-29: an anti-human immunodeficiency virus protein from
RT Trichosanthes kirilowii that is nontoxic to intact cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:6570-6574(1991).
CC -1- FUNCTION: CAPABLE OF INHIBITING HIV-1 INFECTION AND REPLICATION.
CC -1- INACTIVATES EUKARYOTIC 60S RIBOSOMAL SUBUNITS.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of the N-glycosidic bond at one
CC specific adenosine on the 28S rRNA.
CC -1- SIMILARITY: BELONGS TO THE RIBOSOME-INACTIVATING PROTEIN FAMILY.
CC TYPE 1 RIP SUBFAMILY.
CC PIR: A39598; A39598.
DR HSSP: P09989; 1MRJ.
DR InterPro: IPR001574; RIP.
DR Pfam: PF00161; RIP; 1.
DR PROSITE: PS00275; SHIGA_RICIN; PARTIAL.
KM Plant defense: Protein synthesis inhibitor; Hydrolase; Toxin.
FT NON TER 45 45
SQ SEQUENCE 45 AA; 5040 MW; BE1178150C1CF7B7 CRC64;

Query Match 1.5%; Score 6; DB 1; Length 45;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 LSGATS 147
Db 6 LSGATS 11

RESULT 32
ILVD RHOCA STANDARD; PRT; 46 AA.
ID ILVD RHOCA
AC P31874;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dihydroxy-acid dehydratase (EC 4.2.1.9) (DAD) (Fragment).
GN ILVD.
OS Rhodospirillum rubrum (Rhodospirillum rubrum).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillum; Rhodospirillum; Rhodospirillum.
OX NCBI_TaxID=1061;
RN NCBI_TaxID=1061;
RP SEQUENCE FROM N.A.
RX MEDLINE=92307401; PubMed=119375;
RA Kranz R.G., Beckman D.L., Foster-Hartnett D.;
RT "DNA gyrase activities from Rhodospirillum rubrum: analysis of
RT target(s) of coumarins and cloning of the gyrB locus.";
RL FEMS Microbiol. Lett. 72:25-32(1992).
CC -1- CATALYTIC ACTIVITY: 2,3-dihydroxy-3-methylbutanoate = 3-methyl-2-
CC oxobutanoate + H(2)O.
CC -1- COFACTOR: BINDS 1 AFE-4S CLUSTER (POTENTIAL).
CC -1- PATHWAY: Valine and isoleucine biosynthesis; fourth step.

```

CC -!- SIMILARITY: BELONGS TO THE ILVD / EDD FAMILY.  
DR InterPro; IPR000591; ILVD\_EDD\_family.  
DR Pfam; PF00920; ILVD\_EDD; 1.  
DR ProDom; PD002691; ILVD\_EDD\_family; 1.  
DR PROSITE; PS00886; ILVD\_EDD\_1; PARTIAL.  
DR PROSITE; PS00887; ILVD\_EDD\_2; PARTIAL.  
KW Branched-chain amino acid biosynthesis; Lyase; Iron-sulfur; 4Fe-4S.  
FT NON\_TER 1  
FT TER 46  
SQ SEQUENCE 46 AA; 5041 MW; F1BB4CF8A3407737 CRC64;  
  
Query Match 1.5%; Score 6; DB 1; Length 46;  
Best Local Similarity 100.0%; Pred. No. 60; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 280 LAVLFQ 285  
Db 37 LAVLFQ 42  
|||||  
-----  
RESULT 33  
YME2\_THIFE  
ID\_YME2\_THIFE STANDARD; PRT; 55 AA.  
AC P22903;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 5.9 kDa protein in MOBE 3'region (ORF 5).  
OS Thiobacillus ferrooxidans.  
OG Plasmid pTF-FC2.  
OC Bacteria; Proteobacteria; gamma subdivision; Acidithiobacillus.  
OX NCBI\_TaxID=920;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93015664; PubMed=1400173;  
RA Rohrer J., Rawlings D.E.;  
RT "Sequence analysis and characterization of the mobilization region of a broad-host-range plasmid, pTF-FC2, isolated from Thiobacillus ferrooxidans."  
RL J. Bacteriol. 174:6230-6237(1992).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M57717; AAA27394.1; -.  
DR PIR; G43256; G43256.  
DR PIR; S27627; S27627.  
KW Hypothetical protein; Plasmid.  
SQ SEQUENCE 55 AA; 5909 MW; 84D11ABA3399FB32A CRC64;  
  
Query Match 1.5%; Score 6; DB 1; Length 55;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 198 VSENGN 203  
Db 50 VSENGN 55  
|||||  
-----  
RESULT 34  
VGLB\_BPPZA  
ID\_VGLB\_BPPZA STANDARD; PRT; 56 AA.  
AC P06948;  
DT 01-JAN-1988 (Rel. 06, Created)  
DT 01-JAN-1988 (Rel. 06, Last sequence update)  
DT 01-JAN-1988 (Rel. 06, Last annotation update)  
DE Early protein GP1B.  
GN 1B.

OS Bacteriophage PZA.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;  
OC phi-29-like viruses.  
OX NCBI\_TaxID=10757;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86056991; PubMed=3934048;  
RA Paces V., Vitek C., Urbanek P., Hostomsky Z.;  
RT "Nucleotide sequence of the major early region of Bacillus subtilis phage PZA, a close relative of phi 29."  
RL Gene 38:45-56(1985).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M11813; AAA88480.1; -.  
DR PIR; B24528; ERBP1B.  
KW Early protein.  
SQ SEQUENCE 56 AA; 6580 MW; 2B791F6D5EED39B0 CRC64;  
  
Query Match 1.5%; Score 6; DB 1; Length 56;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 294 KGLSLS 299  
Db 29 KGLSLS 34  
|||||  
-----  
RESULT 35  
YSCE\_YEREN  
ID\_YSCE\_YEREN STANDARD; PRT; 66 AA.  
AC Q01246;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 01-APR-1993 (Rel. 25, Last sequence update)  
DT 01-FEB-1995 (Rel. 31, Last annotation update)  
DE Putative YOP proteins translocation protein E.  
GN YSCE.  
OS Yersinia enterocolitica.  
OG Plasmid pYV.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OX NCBI\_TaxID=630;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=439-80 / Serotype O:9;  
RX MEDLINE=91317716; PubMed=1860816;  
RA Michiels T., Vanooteghem J.-C., de Rouvroit C., China B., Gustin A., Boudry P., Cornelis G.R.;  
RT "Analysis of virC, an operon involved in the secretion of Yop proteins by Yersinia enterocolitica."  
RL J. Bacteriol. 173:4994-5009(1991).  
CC -!- INDUCTION: AT 37 DEGREES CELSIUS IN THE ABSENCE OF CALCIUM.  
CC -!- MISCELLANEOUS: BELONGS TO AN OPERON INVOLVED IN THE TRANSLOCATION OF YOP PROTEINS ACROSS THE BACTERIAL MEMBRANES OR IN THE SPECIFIC CONTROL OF THIS FUNCTION.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M74011; AAC37022.1; -.  
DR PIR; B40361; E40361.  
KW Hypothetical protein; Plasmid; Virulence.

SQ SEQUENCE 66 AA; 7449 MW; 40CCFA56F84C863 CRC64;  
 Query Match 1.5%; Score 6; DB 1; Length 66;  
 Best Local Similarity 100.0%; Pred. No. 84;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 378 QSESKA 383  
 |||||  
 DB 45 QSESKA 50

RESULT 36  
 NSGX HUMAN  
 ID NSGX\_HUMAN STANDARD; PRT; 79 AA.  
 AC Q9UH64;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Susceptibility protein NSG-X.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Neopharyngeal carcinoma;  
 RA Jun Q., Yang J.B., Li G.Y.;  
 RT "A new nasopharyngeal carcinoma associated gene on 9p21-22."  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Burian D.M., Mitchell N., Roe B.A.;  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
 DR EMBL; AF211119; AAF22974.1;  
 DR EMBL; AC000048; -; NOT ANNOTATED CDS.  
 SQ SEQUENCE 79 AA; 8851 MW; EC94CB8C34FDF5F8C CRC64;

Query Match 1.5%; Score 6; DB 1; Length 79;  
 Best Local Similarity 100.0%; Pred. No. 99;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 PSIGVS 93  
 |||||  
 DB 48 PSIGVS 53

RESULT 37  
 UR2 PLAFB  
 ID UR2\_PLAFB STANDARD; PRT; 83 AA.  
 AC P21857;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Uroensin II precursor (U-II) (Uii) (Fragments).  
 OS Platyhelminthes (European flounder).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Perciformes; Pleuronectiformes;  
 OC Pleuronectoidae; Pleuronectidae; Platycthyes.  
 OX NCBI\_TaxID=8260;  
 RN [1]  
 RP SEQUENCE.  
 RC TISSUE=Urophysalis;  
 RX MEDLIN=90306357; PubMed=2365069;

RA Conlon J.M., Arnold-Reed D.E., Balmert R.J.;  
 RT "Post-translational processing of prepro-urotensin II.";  
 RL FEBS Lett. 266:37-40(1990).  
 CC -1- FUNCTION: UROSENSIN IS FOUND IN THE TELEOST CAUDAL NEUROSECRETORY  
 SYSTEM. IT HAS A SUGGESTED ROLE IN OSMOREGULATION AND AS A  
 CORTICOTROPIN-RELEASING FACTOR. THE NONHORMONAL PORTION OF THIS  
 PRECURSOR MAY BE A UROSENSIN BINDING PROTEIN, UROPHYSIN.  
 CC -1- SIMILARITY: BELONGS TO THE UROSENSIN 2 FAMILY.  
 DR PIR; S10706; S10706.  
 DR InterPro; IPR001483; Uroensin II.  
 DR PROSITE; PS00984; UROSENSIN II.1.  
 KM Hormone; Cleavage on pair of basic residues.  
 FT NON\_TER 1  
 FT PEPTIDE 1 >48 UROPHYSIN (POTENTIAL).  
 FT NON\_CONS 48 49  
 FT NON\_CONS 71 72  
 FT PEPTIDE 72 83 UROSENSIN II.  
 FT DISULFID 77 82  
 SQ SEQUENCE 83 AA; 9292 MW; 6DD057577BE6DF703 CRC64;

Query Match 1.5%; Score 6; DB 1; Length 83;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 296 LSISEQ 301  
 |||||  
 DB 28 LSISEQ 33

RESULT 38  
 Y16\_BPT7  
 ID Y16\_BPT7 STANDARD; PRT; 86 AA.  
 AC P03793;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Hypothetical gene 1.6 protein.  
 GN 1.6.  
 OS Bacteriophage T7.  
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;  
 OC T7-like viruses.  
 OX NCBI\_TaxID=10760;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLIN=83241725; PubMed=6864790;  
 RA Dunn J.J., Studier F.W.;  
 RT "Complete nucleotide sequence of bacteriophage T7 DNA and the  
 locations of T7 genetic elements.";  
 RL J. Mol. Biol. 166:477-535(1983).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLIN=82078034; PubMed=7310871;  
 RA Dunn J.J., Studier F.W.;  
 RT "Nucleotide sequence from the genetic left end of bacteriophage T7  
 DNA to the beginning of gene 4.";  
 RL J. Mol. Biol. 148:303-330(1981).  
 RN [3]  
 RP SEQUENCE OF 1-18 FROM N.A.  
 RX MEDLIN=80144900; PubMed=231766;  
 RA Boothroyd J.C., Hayward R.S.;  
 RT "New genes and promoters suggested by the DNA sequence near the end  
 of the coliphage T7 early operon.";  
 RL Nucleic Acids Res. 7:1931-1943(1979).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
 DR EMBL; V01146; CAA24396.1; -;

```

DR EMBL; V01127; CAA24339.1; -.
DR EMBL; M25443; AAA32567.1; -.
DR PIR; A04418; Q1BP67.
DR PIR; S42295; S42295.
KW Hypothetical protein.
SQ SEQUENCE 86 AA; 9946 MW; 648D9DCODD12427F CRC64;
Query Match 1.5%; Score 6; DB 1; Length 86;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 15 LSVAVT 20
Db 65 LSVAVT 70

RESULT 39
IHFB_BUCAI
ID IHFB_BUCAI STANDARD; PRT; 94 AA.
AC P57394;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Integration host factor beta-subunit (IHFB-beta).
GN IHFB OR HMD OR BU308.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
Bacteria; Proteobacteria).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Tokyo 1998;
RC MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -1- FUNCTION: THIS PROTEIN IS ONE OF THE TWO SUBUNITS OF INTEGRATION
CC -1- HOST FACTOR, A SPECIFIC DNA-BINDING PROTEIN THAT FUNCTIONS IN
CC GENETIC RECOMBINATION AS WELL AS IN TRANSCRIPTIONAL AND
CC TRANSCRIPTIONAL CONTROL (BY SIMILARITY).
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP001118; BAB13017.1; -.
CC HSSP; P08756; 1LHF.
DR InterPro; IPR000119; Bac_DNABind.
DR Pfam; PF00216; Bac_DNA_binding; 1.
DR ProDom; PD000945; Bac_DNABind; 1.
DR SMART; SMO0411; BHL; 1.
DR TIGRFAMs; TIGR00988; hfp; 1.
DR PROSITE; PS00045; HISTONE_LIKE; 1.
KW DNA-binding; Transcription regulation; DNA recombination;
KW Translation regulation; Complete proteome.
SQ SEQUENCE 94 AA; 10974 MW; 1BA9CC3084734CA CRC64;
Query Match 1.5%; Score 6; DB 1; Length 94;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 221 ANGKRI 226
Db 38 ANGKRI 43

```

```

RESULT 40
YDFF_SCHPO
ID YDFF_SCHPO STANDARD; PRT; 94 AA.
AC Q10486;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein C17C9.15c in chromosome I.
GN SPAC17C9.15C.
OS Schizosaccharomyces pombe (Fission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voickaert G., Aert R., Robben J., Grymonprez B.,
RA Welter J., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mortier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Bartell B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z73099; CAA97345.1; -.
DR Hypothetical protein.
KW
SQ SEQUENCE 94 AA; 11248 MW; 3CECAC60D5C962E0 CRC64;
Query Match 1.5%; Score 6; DB 1; Length 94;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 164 SDLEE 169
Db 85 SDLEE 90

RESULT 41
Y11K_BPCHP
ID Y11K_BPCHP STANDARD; PRT; 96 AA.
AC P19183; P19190;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)

```

DE Hypothetical 11.5 kDa protein (ORF5).  
 OS Bacteriophage Chp1.  
 OC Viruses; ssDNA viruses; Microviridae; Microvirus.  
 OX NCBI\_TaxID=12367;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90111716; PubMed=2607341;  
 RA Storey C.C., Lusher M., Richmond S.J.;  
 RT "Analysis of the complete nucleotide sequence of Chp1, a phage which  
 RT infects avian Chlamydia psittaci."  
 RL J. Gen. Virol. 70:3381-3390(1989).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; D00624; BAA00513.1; -  
 CC EMBL; D00624; BAA00514.1; ALT\_INIT.  
 CC PIR; J00349; J00349.  
 CC Hypothetical protein.  
 KM SEQUENCE 96 AA; 11469 MW; 466C622144B59323 CRC64;  
 SQ

Query Match 1.5%; Score 6; DB 1; Length 96;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 333 KGRFYP 338  
 |||||

DB 66 KGRFYP 71

RESULT 42  
 YACB RHISN STANDARD; PRT; 98 AA.  
 AC P55384;  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 GN Hypothetical 10.2 kDa protein YACB.  
 OS YACB.  
 OS Rhizobium sp. (strain NGR234).  
 OC Plasmid sym pNGR234a.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 OX NCBI\_TaxID=394;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97305956; PubMed=9163424;  
 RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,  
 RA Perret X.;  
 RT "Molecular basis of symbiosis between Rhizobium and legumes."  
 RL Nature 387:394-401(1997).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
 CC -1- SIMILARITY: NONE OBVIOUS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; A800067; AAB91632.1; -  
 CC Hypothetical protein; Plasmid; Transmembrane.  
 KM TRANSMEM 2 22 POTENTIAL.  
 FT TRANSMEM 70 90  
 SQ SEQUENCE 98 AA; 10241 MW; B373326450613B5 CRC64;  
 RX

Query Match 1.5%; Score 6; DB 1; Length 98;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 348 ALPILQ 353  
 |||||

DB 86 ALPILQ 91

RESULT 43  
 YUN5\_YEAST STANDARD; PRT; 105 AA.  
 ID YUN5\_YEAST  
 AC P47012;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Hypothetical 12.4 kDa protein in RPS21B-MRS3 intergenic region.  
 GN YUL135W OR J0666.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=8286C / FY1679;  
 RX MEDLINE=96408771; PubMed=8813765;  
 RA Katsoulou C., Tzeremia M., Tavernarakis N., Alexandraki D.;  
 RT "Sequence analysis of a 40.7 kb segment from the left arm of yeast  
 RT chromosome X reveals 14 known genes and 13 new open reading frames  
 RT including homologues of genes clustered on the right arm of  
 RT chromosome XI."  
 RL Yeast 12:787-797(1996).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X87371; CAA60820.1; -  
 CC EMBL; 249410; CAA89429.1; -  
 CC SGB; S0003671; YUL135W.  
 KM Hypothetical protein.  
 SQ SEQUENCE 105 AA; 12374 MW; DE09817D05033D6C CRC64;  
 RX

Query Match 1.5%; Score 6; DB 1; Length 105;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 RLAVLP 284  
 |||||

DB 48 RLAVLP 53

RESULT 44  
 SS18\_STRVG STANDARD; PRT; 111 AA.  
 ID SS18\_STRVG  
 AC P80388;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Subtilisin inhibitor-like protein-8 (SIL-8) (SIL8).  
 GN SIL8.  
 OS Streptomyces virginiae.  
 OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;  
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1961;  
 RN [1]  
 RP SEQUENCE.  
 RC STRAIN=4425;  
 RX MEDLINE=95094822; PubMed=8001578;

RA Terabe M., Kojima S., Taguchi S., Momose H., Miura K.-I.;  
RT "Primary structure and inhibitory properties of a subtilisin-  
RL chymotrypsin inhibitor from *Streptomyces virginiae*.";  
RL Eur. J. Biochem. 226:627-632(1994).  
CC -1- FUNCTION: INHIBITOR OF SUBTILISIN AND CHYMOTRYPSIN.  
CC -1- SUBUNIT: HOMODIMER.  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- SIMILARITY: BELONGS TO THE SERINE PROTEASE INHIBITORS SSI FAMILY.  
DR HSSP; P01006; 3SSI.  
DR InterPro: IPR000691; Strep\_subst\_inhib.  
DR Pfam: PF00720; SSI; 1.  
DR PRINTS; PR00294; SSETLINHETR.  
DR PRODOM; PD04028; Strep\_subst\_inhib; 1.  
DR PROSITE; PS00999; SSI; 1.  
FW Serine protease inhibitor.  
KT DISULFID 31 46 BY SIMILARITY.  
FT DISULFID 69 99 BY SIMILARITY.  
FT ACT\_SITE 71 72 REACTIVE BOND.  
SQ SEQUENCE 111 AA; 11436 MW; 02F6F060A04E8530 CRC64;  
  
Query Match 1.5%; Score 6; DB 1; Length 111;  
Best Local Similarity 100.0%; Pred.No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 215 TGGAFD 220  
Db 52 TGGAFD 57  
  
RESULT 45  
YCX5 OENHO  
ID YCX5 OENHO STANDARD; PRT; 111 AA.  
AC Q9MTN3;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Hypothetical 12.8 kDa protein in ycf9-trns intergenic region (ORF11).  
OS *Oenothera hookeri* (Hooker's evening primrose).  
OG Chloroplast.  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;  
OC eurosids II; Myrtales; Onagraceae; Oenothera.  
OX NCBI\_TaxID=85636;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Johansen;  
RX MEDLINE=20309318; PubMed=10852478;  
RA Hupfer H., Swiatek M., Hornung S., Herrmann R.G., Maier R.M.,  
RA Chiu W.-L., Sears B.;  
RT "Complete nucleotide sequence of the *Oenothera elata* plastid  
RT chromosome, representing plastome I of the five distinguishable  
RT *Oenothera* plastomes.";  
RL Mol. Gen. Genet. 263:581-585(2000).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AJ271079; CAB67142.1; -.  
KW Chloroplast; Hypothetical protein.  
SQ SEQUENCE 111 AA; 12814 MW; E5E0CE989317F140 CRC64;  
  
Query Match 1.5%; Score 6; DB 1; Length 111;  
Best Local Similarity 100.0%; Pred.No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 85 QLIPSL 90  
Db 82 QLIPSL 87

Search completed: December 25, 2002, 20:31:08  
Job time : 35 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Comugen Ltd.

OM protein - protein search, using sw model

Run on: December 25, 2002, 20:28:03 ; Search time 31 Seconds  
(without alignments)  
2658.672 Million cell updates/sec

Title: US-09-889-746-2

Perfect score: 400

Sequence: 1 MHSHYFQWLSPLSVAVT.....SKAEVLGRVPLNKPRALF 400

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

SPTREMBL\_21:\*  
1: sp\_archea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacterioplasmid:\*  
17: sp\_archaeal:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13	3.2	705	2	Q9ZNM3
2	13	3.2	725	2	Q93405
3	10	2.5	726	16	Q8ZHB0
4	10	2.5	732	2	Q9XCH0
5	10	2.5	732	2	Q93PE2
6	9	2.2	813	16	Q913X9
7	8	2.0	334	16	Q86513
8	8	2.0	335	17	Q87293
9	8	2.0	812	16	Q8X5Q1
10	8	2.0	1561	4	Q9UL61
11	8	2.0	2621	5	Q8SSW5
12	7	1.8	63	16	Q8ZST7
13	7	1.8	68	4	Q9Y3H9
14	7	1.8	79	2	Q9X8G0
15	7	1.8	83	13	Q9YCX0
16	7	1.8	83	13	Q9PWJ4

17	7	1.8	83	13	Q9YGC7	Q9YGC7 laticauda c
18	7	1.8	83	13	Q9PRJ7	Q9PRJ7 laticauda c
19	7	1.8	83	13	Q9PRJ6	Q9PRJ6 laticauda c
20	7	1.8	83	13	Q9PRJ5	Q9PRJ5 laticauda c
21	7	1.8	83	13	Q9PRJ0	Q9PRJ0 laticauda c
22	7	1.8	85	16	Q8VX3	Q8VX3 mycobacteri
23	7	1.8	87	16	Q9CUE2	Q9CUE2 lactococcus
24	7	1.8	95	10	Q9LMD0	Q9LMD0 oryza sativ
25	7	1.8	101	10	Q9M581	Q9M581 spiniacia ol
26	7	1.8	108	11	Q9DBM6	Q9DBM6 mus musculu
27	7	1.8	108	16	Q8ZNU2	Q8ZNU2 salmonella
28	7	1.8	115	12	Q9PXU5	Q9PXU5 sugarcane m
29	7	1.8	130	8	Q9MH59	Q9MH59 ginkgo bilo
30	7	1.8	154	2	Q45435	Q45435 bacillus sp
31	7	1.8	156	12	Q8ORX5	Q8ORX5 chimpanzee
32	7	1.8	161	13	Q90494	Q90494 eptaretus
33	7	1.8	173	17	Q9HNS0	Q9HNS0 halobacteri
34	7	1.8	182	16	Q9F3Q2	Q9F3Q2 streptomyce
35	7	1.8	183	5	Q9VR84	Q9VR84 dirosophila
36	7	1.8	187	11	Q9QZJ2	Q9QZJ2 mesocricetu
37	7	1.8	188	5	Q61000	Q61000 crithidia f
38	7	1.8	188	5	Q9RZX2	Q9RZX2 crithidia f
39	7	1.8	190	5	Q9GQW7	Q9GQW7 leishmania
40	7	1.8	190	5	Q9SWZ8	Q9SWZ8 leishmania
41	7	1.8	199	5	Q9BP39	Q9BP39 leishmania
42	7	1.8	199	5	Q95WL4	Q95WL4 leishmania
43	7	1.8	199	5	Q95NFS	Q95NFS leishmania
44	7	1.8	200	16	Q9HVB2	Q9HVB2 pseudomonas
45	7	1.8	202	12	Q8VSK7	Q8VSK7 possum aden
46	7	1.8	210	10	Q9FJ14	Q9FJ14 arabidopsis
47	7	1.8	213	5	Q9W0N7	Q9W0N7 dirosophila
48	7	1.8	213	5	Q9WC46	Q9WC46 bacterioph
49	7	1.8	213	13	Q9DFQ2	Q9DFQ2 gliichivys
50	7	1.8	223	16	Q8VLO1	Q8VLO1 anabena sp
51	7	1.8	225	16	Q9X8T6	Q9X8T6 streptomyce
52	7	1.8	227	5	Q8SR25	Q8SR25 encaphalito
53	7	1.8	230	5	Q9N9X2	Q9N9X2 geodia cydo
54	7	1.8	230	11	Q9D1R1	Q9D1R1 mus musculu
55	7	1.8	233	16	Q9PFR7	Q9PFR7 xyella fas
56	7	1.8	248	11	Q9D423	Q9D423 mus musculu
57	7	1.8	249	16	Q9KDL9	Q9KDL9 bacillus ha
58	7	1.8	252	11	Q9CKS4	Q9CKS4 mus musculu
59	7	1.8	254	17	Q97VW5	Q97VW5 sulfolobus
60	7	1.8	259	5	Q8SSN8	Q8SSN8 dictyosteli
61	7	1.8	260	5	Q76669	Q76669 caenorhabdi
62	7	1.8	270	8	Q48257	Q48257 demmyus hir
63	7	1.8	271	16	Q8U7G3	Q8U7G3 agrobacteri
64	7	1.8	275	17	Q27290	Q27290 methanobact
65	7	1.8	276	17	Q96XS1	Q96XS1 sulfolobus
66	7	1.8	281	10	Q9LMT2	Q9LMT2 arabidopsis
67	7	1.8	282	12	Q91CG3	Q91CG3 kaposi's sa
68	7	1.8	289	16	Q9CMN9	Q9CMN9 pasteurella
69	7	1.8	290	16	Q8X4E8	Q8X4E8 escherichia
70	7	1.8	293	16	Q9W9J1	Q9W9J1 staphylococ
71	7	1.8	303	16	Q8XUA8	Q8XUA8 ralsstonia s
72	7	1.8	321	2	Q07J32	Q07J32 rhizobium m
73	7	1.8	325	16	Q8YB90	Q8YB90 bruceella me
74	7	1.8	333	11	Q9BPZ9	Q9BPZ9 mus musculu
75	7	1.8	333	11	Q924K0	Q924K0 mus musculu
76	7	1.8	334	16	Q9CKG7	Q9CKG7 pasteurella
77	7	1.8	334	16	Q92249	Q92249 rhizobium m
78	7	1.8	335	16	Q8UBV6	Q8UBV6 agrobacteri
79	7	1.8	337	16	Q9KRW5	Q9KRW5 vibrio chol
80	7	1.8	340	16	Q9VBL1	Q9VBL1 deinococcus
81	7	1.8	348	12	Q9BMY0	Q9BMY0 anabena mo
82	7	1.8	349	16	Q8YHV2	Q8YHV2 bruceella me
83	7	1.8	351	16	Q8ZCB6	Q8ZCB6 yersinia pe
84	7	1.8	354	16	Q97N02	Q97N02 clostridium
85	7	1.8	361	16	Q51657	Q51657 borrelia bu
86	7	1.8	362	16	Q8YXA2	Q8YXA2 anabena sp
87	7	1.8	369	16	Q8X928	Q8X928 escherichia
88	7	1.8	376	16	Q8ZRF6	Q8ZRF6 salmonella
89	7	1.8	385	16	Q8Z829	Q8Z829 salmonella

90	7	1.8	386	16	Q9ZK50	Q9ZK50 helicobacte	163	7	1.8	747	10	O82007	O82007 lycopersico
91	7	1.8	386	16	O25780	Q25780 helicobacte	164	7	1.8	747	10	Q9SAN2	Q9SAN2 lycopersico
92	7	1.8	390	15	Q8ZBA9	Q8ZBA9 yersinia pe	165	7	1.8	754	10	O65834	O65834 lycopersico
93	7	1.8	395	3	Q96UC5	Q96UC5 yersinia	166	7	1.8	754	10	O65835	O65835 lycopersico
94	7	1.8	396	2	Q9ZJ20	Q9ZJ20 lactococcus	167	7	1.8	803	16	Q9KNP7	Q9KNP7 vibrio chol
95	7	1.8	396	2	O60338	O60338 lactococcus	168	7	1.8	804	12	Q98562	Q98562 paramedium
96	7	1.8	398	16	Q8XH75	Q8XH75 clostridium	169	7	1.8	815	4	O43273	O43273 homo sapien
97	7	1.8	412	4	Q9NWV1	Q9NWV1 homo sapien	170	7	1.8	821	10	Q8VZJ4	Q8VZJ4 arabisdopsis
98	7	1.8	416	10	Q9M7Q4	Q9M7Q4 arabisdopsis	171	7	1.8	821	10	Q9SW10	Q9SW10 arabisdopsis
99	7	1.8	418	2	Q9F413	Q9F413 mycoplasma	172	7	1.8	867	6	Q8WN91	Q8WN91 bos taurus
100	7	1.8	424	10	Q9FGM0	Q9FGM0 arabisdopsis	173	7	1.8	867	16	Q8VTP0	Q8VTP0 onabaeana sp
101	7	1.8	425	15	Q9A454	Q9A454 caulobacter	174	7	1.8	931	6	Q9BDW2	Q9BDW2 chaetophrac
102	7	1.8	430	16	Q9WZT4	Q9WZT4 thermotoga	175	7	1.8	980	4	Q9NS55	Q9NS55 homo sapien
103	7	1.8	432	16	Q8XK98	Q8XK98 clostridium	176	7	1.8	1029	10	O23015	O23015 arabisdopsis
104	7	1.8	433	16	Q8ZRY4	Q8ZRY4 salmonella	177	7	1.8	1032	16	Q9KYL1	Q9KYL1 streptomyce
105	7	1.8	433	16	Q8ZQ56	Q8ZQ56 salmonella	178	7	1.8	1033	11	Q920L3	Q920L3 mus musculus
106	7	1.8	433	16	Q8Z3E5	Q8Z3E5 salmonella	179	7	1.8	1045	4	Q9UNR9	Q9UNR9 homo sapien
107	7	1.8	436	2	Q9AMX7	Q9AMX7 bradyrhizob	180	7	1.8	1045	4	Q9NS56	Q9NS56 homo sapien
108	7	1.8	437	16	Q8XMU8	Q8XMU8 clostridium	181	7	1.8	1049	2	Q9ZEM4	Q9ZEM4 streptomyce
109	7	1.8	447	11	Q920C1	Q920C1 mus musculus	182	7	1.8	1051	4	Q9H7M6	Q9H7M6 homo sapien
110	7	1.8	450	4	Q9BU40	Q9BU40 homo sapien	183	7	1.8	1051	10	O49494	O49494 arabisdopsis
111	7	1.8	453	16	O66852	O66852 aquifex aeo	184	7	1.8	1054	11	Q8VCD7	Q8VCD7 mus musculus
112	7	1.8	454	12	Q86339	Q86339 reovirus sp	185	7	1.8	1089	5	O26155	O26155 plasmodium
113	7	1.8	454	12	Q86340	Q86340 reovirus sp	186	7	1.8	1110	2	Q8RL73	Q8RL73 pseudomonas
114	7	1.8	456	13	Q90ZD5	Q90ZD5 gallus gall	187	7	1.8	1125	16	Q9A3K9	Q9A3K9 caulobacter
115	7	1.8	462	12	Q85663	Q85663 reovirus (t	188	7	1.8	1153	10	O850G6	O850G6 cryza nativ
116	7	1.8	466	10	O23244	O23244 arabisdopsis	189	7	1.8	1206	11	O08815	O08815 rattus norv
117	7	1.8	466	16	Q8ZF77	Q8ZF77 yersinia pe	190	7	1.8	1215	16	P96937	P96937 mycobacteri
118	7	1.8	469	12	Q9WB52	Q9WB52 physalis se	191	7	1.8	1234	2	Q9AES2	Q9AES2 clostridium
119	7	1.8	469	12	Q91QP5	Q91QP5 melon yell	192	7	1.8	1267	5	Q9U0U5	Q9U0U5 leishmania
120	7	1.8	486	12	Q82042	Q82042 human rotav	193	7	1.8	1381	16	O53552	O53552 mycobacteri
121	7	1.8	486	12	O40626	O40626 human rotav	194	7	1.8	1384	16	Q8VIZ1	Q8VIZ1 mycobacteri
122	7	1.8	491	2	O88107	O88107 vibrio para	195	7	1.8	1398	16	Q8VKH0	Q8VKH0 mycobacteri
123	7	1.8	500	16	Q8Y065	Q8Y065 ralstonia s	196	7	1.8	1416	5	Q9W539	Q9W539 dirosophila
124	7	1.8	508	16	Q8X314	Q8X314 escherichia	197	7	1.8	1676	16	Q9KXX9	Q9KXX9 streptomyce
125	7	1.8	519	10	Q9LYN9	Q9LYN9 arabisdopsis	198	7	1.8	1854	2	O33759	O33759 streptococc
126	7	1.8	527	5	Q93762	Q93762 caenorhabdi	199	7	1.8	1854	2	O33760	O33760 streptococc
127	7	1.8	535	2	Q9RW66	Q9RW66 pseudomonas	200	7	1.8	1854	2	O33764	O33764 streptococc
128	7	1.8	535	10	Q9AVB8	Q9AVB8 cryza sativ	201	7	1.8	1874	2	O33765	O33765 streptococc
129	7	1.8	537	2	O33417	O33417 pseudomonas	202	7	1.8	1995	3	O14029	O14029 schizosacch
130	7	1.8	538	16	Q916U8	Q916U8 pseudomonas	203	7	1.8	2439	5	Q9VMS2	Q9VMS2 dirosophila
131	7	1.8	542	16	Q9ZMS3	Q9ZMS3 rhizobium m	204	7	1.8	2785	17	Q8ZYB9	Q8ZYB9 pyrobaculum
132	7	1.8	546	16	Q9ZMD7	Q9ZMD7 helicobacte	205	7	1.8	3109	5	Q9BMQ0	Q9BMQ0 dirosophila
133	7	1.8	563	12	Q8V6K1	Q8V6K1 halovirus h	206	7	1.8	3659	16	Q9BLN6	Q9BLN6 rhizobium l
134	7	1.8	563	16	Q92WS0	Q92WS0 rhizobium m	207	7	1.8	4835	5	O8T5T1	O8T5T1 giardia lam
135	7	1.8	566	5	Q9U0J5	Q9U0J5 plasmodium	208	6	1.5	27	12	Q9Q129	Q9Q129 hepatitis c
136	7	1.8	580	16	Q9KBX1	Q9KBX1 bacillus ha	209	6	1.5	27	12	Q9Q128	Q9Q128 hepatitis c
137	7	1.8	587	15	O25510	O25510 helicobacte	210	6	1.5	27	12	Q9Q122	Q9Q122 hepatitis c
138	7	1.8	587	16	Q9ZL05	Q9ZL05 helicobacte	211	6	1.5	27	12	Q9Q118	Q9Q118 hepatitis c
139	7	1.8	590	2	Q45819	Q45819 clostridium	212	6	1.5	27	12	Q9Q116	Q9Q116 hepatitis c
140	7	1.8	613	2	Q93HH7	Q93HH7 streptomyce	213	6	1.5	27	12	Q9Q115	Q9Q115 hepatitis c
141	7	1.8	637	10	O8SAD9	O8SAD9 triticum mo	214	6	1.5	27	12	Q9Q113	Q9Q113 hepatitis c
142	7	1.8	641	10	Q9FNU2	Q9FNU2 cryza sativ	215	6	1.5	27	12	Q91KX8	Q91KX8 hepatitis c
143	7	1.8	644	10	Q9ARL6	Q9ARL6 hordeum vul	216	6	1.5	27	12	Q91KX6	Q91KX6 hepatitis c
144	7	1.8	644	10	Q9ARL6	Q9ARL6 hordeum vul	217	6	1.5	27	12	Q91KX5	Q91KX5 hepatitis c
145	7	1.8	644	10	O8SAE0	O8SAE0 triticum mo	218	6	1.5	27	12	Q91KX4	Q91KX4 hepatitis c
146	7	1.8	648	16	Q8X9H9	Q8X9H9 escherichia	219	6	1.5	27	12	Q91KX3	Q91KX3 hepatitis c
147	7	1.8	662	16	Q8Y2D6	Q8Y2D6 ralstonia s	220	6	1.5	27	12	Q91KX2	Q91KX2 hepatitis c
148	7	1.8	665	5	O18597	O18597 caenorhabdi	221	6	1.5	27	12	Q91KX1	Q91KX1 hepatitis c
149	7	1.8	666	10	Q9ZRA6	Q9ZRA6 lycopersico	222	6	1.5	27	12	Q91KX0	Q91KX0 hepatitis c
150	7	1.8	668	16	Q99X13	Q99X13 staphylococ	223	6	1.5	27	12	Q91KX0	Q91KX0 hepatitis c
151	7	1.8	674	10	Q9LZ86	Q9LZ86 arabisdopsis	224	6	1.5	27	12	Q91KW9	Q91KW9 hepatitis c
152	7	1.8	684	10	O8SG67	O8SG67 cryza sativ	225	6	1.5	27	12	Q91KW7	Q91KW7 hepatitis c
153	7	1.8	696	5	O17479	O17479 hyalophora	226	6	1.5	27	12	Q91KW5	Q91KW5 hepatitis c
154	7	1.8	704	10	Q93VF5	Q93VF5 nicotiana t	227	6	1.5	27	12	Q91KW4	Q91KW4 hepatitis c
155	7	1.8	717	16	O8UHN6	O8UHN6 agrobacteri	228	6	1.5	27	12	Q91KW3	Q91KW3 hepatitis c
156	7	1.8	726	2	O85995	O85995 spingomona	229	6	1.5	27	12	Q91KW2	Q91KW2 hepatitis c
157	7	1.8	743	10	Q9LWA4	Q9LWA4 lycopersico	230	6	1.5	27	12	Q91KW1	Q91KW1 hepatitis c
158	7	1.8	745	10	O04678	O04678 lycopersico	231	6	1.5	27	12	Q91KW0	Q91KW0 hepatitis c
159	7	1.8	745	10	O96478	O96478 lycopersico	232	6	1.5	27	12	Q91KV9	Q91KV9 hepatitis c
160	7	1.8	746	6	Q9BE82	Q9BE82 macaca fasc	233	6	1.5	27	12	Q91KV8	Q91KV8 hepatitis c
161	7	1.8	746	10	O9LWA3	O9LWA3 lycopersico	234	6	1.5	27	12	Q91KV7	Q91KV7 hepatitis c
162	7	1.8	747	10	O65836	O65836 lycopersico	235	6	1.5	27	12	Q91KV6	Q91KV6 hepatitis c

236	6	1.5	27	12	Q91KV5	Q91kv5 hepatitis c	309	6	1.5	79	11	Q9QU0	Q9qu0 mus musculus
237	6	1.5	27	12	Q91KV4	Q91kv4 hepatitis c	310	6	1.5	79	12	Q9J376	Q9j376 hepatitis c
238	6	1.5	27	12	Q91KV2	Q91kv2 hepatitis c	311	6	1.5	80	12	Q9WN73	Q9wn73 hepatitis c
239	6	1.5	27	12	Q91KV1	Q91kv1 hepatitis c	312	6	1.5	84	5	Q16798	Q16798 paracentrot
240	6	1.5	27	12	Q91KV9	Q91kv9 hepatitis c	313	6	1.5	84	5	Q16803	Q16803 paracentrot
241	6	1.5	27	12	Q91KV8	Q91kv8 hepatitis c	314	6	1.5	84	12	Q9J2A9	Q9j2a9 hepatitis c
242	6	1.5	27	12	Q91KV7	Q91kv7 hepatitis c	315	6	1.5	84	15	Q9J930	Q9j930 human immun
243	6	1.5	27	12	Q91KV6	Q91kv6 hepatitis c	316	6	1.5	85	2	Q9J7B4	Q9j7b4 pseudomonas
244	6	1.5	27	12	Q9J3M0	Q9j3m0 hepatitis c	317	6	1.5	85	5	Q96709	Q96709 illyanassa o
245	6	1.5	27	12	Q9J3L9	Q9j3l9 hepatitis c	318	6	1.5	86	8	Q9XP44	Q9xp44 nebrisa geb1
246	6	1.5	27	12	Q9J3I8	Q9j3i8 hepatitis c	319	6	1.5	86	13	Q98TV6	Q98tv6 salmo galar
247	6	1.5	27	12	Q9J3I7	Q9j3i7 hepatitis c	320	6	1.5	87	12	Q88293	Q88293 small round
248	6	1.5	27	12	Q9J3I6	Q9j3i6 hepatitis c	321	6	1.5	87	15	Q9EAK7	Q9eak7 human immun
249	6	1.5	27	12	Q9J3I5	Q9j3i5 hepatitis c	322	6	1.5	87	16	Q8YIQ3	Q8yiq3 bruceella me
250	6	1.5	27	12	Q9J3I4	Q9j3i4 hepatitis c	323	6	1.5	88	8	Q9G2E5	Q9g2e5 eretmoceru
251	6	1.5	27	12	Q9J3I3	Q9j3i3 hepatitis c	324	6	1.5	88	10	Q94HY1	Q94hy1 oryza sativ
252	6	1.5	27	12	Q9J3I2	Q9j3i2 hepatitis c	325	6	1.5	88	12	Q9J3A2	Q9j3a2 hepatitis c
253	6	1.5	27	12	Q9J3I1	Q9j3i1 hepatitis c	326	6	1.5	89	12	Q8QJ23	Q8qj23 camelpox vi
254	6	1.5	27	12	Q9J3L1	Q9j3l1 hepatitis c	327	6	1.5	89	15	Q9QPM1	Q9qpm1 human immun
255	6	1.5	27	12	Q9J3L0	Q9j3l0 hepatitis c	328	6	1.5	91	8	Q9WR84	Q9wr84 chlorarachn
256	6	1.5	27	12	Q9J3K9	Q9j3k9 hepatitis c	329	6	1.5	91	12	Q9W8O7	Q9w8o7 hepatitis c
257	6	1.5	27	12	Q9J3K8	Q9j3k8 hepatitis c	330	6	1.5	91	12	Q9W8B1	Q9w8b1 hepatitis c
258	6	1.5	27	12	Q9J3K7	Q9j3k7 hepatitis c	331	6	1.5	91	12	Q9WN44	Q9wn44 hepatitis c
259	6	1.5	27	12	Q9J3K6	Q9j3k6 hepatitis c	332	6	1.5	91	12	Q9WN43	Q9wn43 hepatitis c
260	6	1.5	27	12	Q9J3K5	Q9j3k5 hepatitis c	333	6	1.5	91	12	Q9WN42	Q9wn42 hepatitis c
261	6	1.5	27	12	Q9J3K4	Q9j3k4 hepatitis c	334	6	1.5	91	12	Q9WN41	Q9wn41 hepatitis c
262	6	1.5	27	12	Q9J3K3	Q9j3k3 hepatitis c	335	6	1.5	91	12	Q9WN40	Q9wn40 hepatitis c
263	6	1.5	27	12	Q9J3K2	Q9j3k2 hepatitis c	336	6	1.5	91	12	Q9WN30	Q9wn30 hepatitis c
264	6	1.5	27	12	Q9J3K1	Q9j3k1 hepatitis c	337	6	1.5	91	12	Q9WN24	Q9wn24 hepatitis c
265	6	1.5	27	12	Q9J3K0	Q9j3k0 hepatitis c	338	6	1.5	91	16	Q8XC16	Q8xc16 escherichia
266	6	1.5	36	16	Q9J1V7	Q9j1v7 staphylococ	339	6	1.5	92	2	Q9F1H3	Q9f1h3 enterococu
267	6	1.5	43	4	Q9N269	Q9n269 homo sapien	340	6	1.5	92	4	Q9UBS1	Q9ubsl homo sapien
268	6	1.5	46	16	Q9N267	Q9n267 staphylococ	341	6	1.5	92	16	Q9CC10	Q9cc10 mycobacteri
269	6	1.5	47	12	Q9N269	Q9n269 hepatitis c	342	6	1.5	93	4	Q9UBK5	Q9ubk5 homo sapien
270	6	1.5	52	4	Q9P1N5	Q9p1n5 homo sapien	343	6	1.5	93	6	Q8SPM0	Q8spm0 canis famli
271	6	1.5	53	16	Q92ZQ2	Q92zq2 rhizobium m	344	6	1.5	94	5	Q9NUU1	Q9nul1 leishmania
272	6	1.5	53	16	Q92D82	Q92d82 listeria in	345	6	1.5	94	16	Q9P065	Q9p065 heliantus
273	6	1.5	53	16	Q8Y8G7	Q8y8g7 listeria mo	346	6	1.5	94	16	Q99TK1	Q99tk1 staphylococ
274	6	1.5	55	5	Q16799	Q16799 paracentrot	347	6	1.5	95	10	Q94J44	Q94j44 oryza sativ
275	6	1.5	55	16	Q914G0	Q914g0 pseudomonas	348	6	1.5	95	16	Q8Y8Z4	Q8y8z4 listeria mo
276	6	1.5	56	9	Q38503	Q38503 bacterioph	349	6	1.5	97	4	Q9NPK2	Q9npk2 homo sapien
277	6	1.5	56	12	Q84719	Q84719 prospect hi	350	6	1.5	97	4	Q8WV56	Q8wv56 homo sapien
278	6	1.5	56	12	Q84721	Q84721 prospect hi	351	6	1.5	97	8	Q9MM50	Q9mm50 bombyx mori
279	6	1.5	56	12	Q84722	Q84722 prospect hi	352	6	1.5	97	16	Q98173	Q98173 rhizobium l
280	6	1.5	56	12	Q84723	Q84723 prospect hi	353	6	1.5	97	16	Q97MJ7	Q97mj7 clostridium
281	6	1.5	56	12	Q84725	Q84725 prospect hi	354	6	1.5	98	6	Q8SPJ6	Q8spj6 sus scrofa
282	6	1.5	56	12	Q84726	Q84726 prospect hi	355	6	1.5	98	17	Q8UOV7	Q8uov7 pyrococcus
283	6	1.5	58	2	Q84728	Q84728 prospect hi	356	6	1.5	100	11	Q9JUH3	Q9juh3 mus musculu
284	6	1.5	58	2	Q8RKZ3	Q8rkz3 providencia	357	6	1.5	100	12	Q9J1L4	Q9j1l4 white spot
285	6	1.5	58	11	Q9ER87	Q9er87 mus musculu	358	6	1.5	100	15	Q88148	Q88148 chimpanzee
286	6	1.5	61	10	Q9CBM4	Q9cbm4 arabidopsis	359	6	1.5	100	16	Q8UHT4	Q8uht4 agrobacteri
287	6	1.5	62	12	Q8VAX8	Q8vax8 white spot	360	6	1.5	101	2	Q9J540	Q9j540 bacillus sp
288	6	1.5	62	16	Q9Z8Y0	Q9z8y0 chlamydia p	361	6	1.5	101	15	Q71968	Q71968 human immun
289	6	1.5	64	5	Q9V0U5	Q9v0u5 drosophila	362	6	1.5	101	15	Q71973	Q71973 human immun
290	6	1.5	66	16	Q9R2Q7	Q9r2q7 pseudomonas	363	6	1.5	102	10	Q9XJ11	Q9xj11 oryza sativ
291	6	1.5	67	2	Q93KS9	Q93ks9 yersinia en	364	6	1.5	103	12	Q8VAZ8	Q8vaz8 white spot
292	6	1.5	68	12	Q91FX4	Q91fx4 chilo iride	365	6	1.5	104	11	Q9D044	Q9d044 mus musculu
293	6	1.5	68	12	Q9PXU6	Q9pxu6 sugarcane m	366	6	1.5	105	4	Q96FY0	Q96fy0 homo sapien
294	6	1.5	69	2	Q93CG0	Q93cg0 staphylococ	367	6	1.5	105	8	Q8WAO6	Q8wao6 elenchus ja
295	6	1.5	69	11	Q8VHP0	Q8vhp0 mus musculu	368	6	1.5	105	16	Q984C8	Q984c8 rhizobium l
296	6	1.5	70	2	Q9RB13	Q9rb13 pectobacter	369	6	1.5	106	2	Q9EYH8	Q9eyh8 yersinia ps
297	6	1.5	70	10	Q9ASN7	Q9asn7 oryza sativ	370	6	1.5	107	15	Q9O593	Q9o593 human immun
298	6	1.5	70	10	Q947L7	Q947l7 beta vulgar	371	6	1.5	109	2	Q49054	Q49054 mycoplasma
299	6	1.5	71	4	Q8RETS	Q8rets homo sapien	372	6	1.5	109	6	Q18893	Q18893 atelies belz
300	6	1.5	71	11	Q9RIE7	Q9rie7 mus musculu	373	6	1.5	109	10	Q9FUJ2	Q9fuj2 spinacia ol
301	6	1.5	71	16	Q946P1	Q946p1 caulobacter	374	6	1.5	109	16	Q8U6Z9	Q8u6z9 agrobacteri
302	6	1.5	74	11	Q91XP7	Q91xp7 rattus norv	375	6	1.5	109	17	Q57992	Q57992 pyrococcus
303	6	1.5	74	13	Q90798	Q90798 gallus gall	376	6	1.5	110	14	Q91R55	Q91r55 uncultured
304	6	1.5	76	9	Q64342	Q64342 bacterioph	377	6	1.5	111	8	Q9WTK3	Q9wtk3 oenothera h
305	6	1.5	79	5	Q9BKX6	Q9bkx6 halictis as	378	6	1.5	111	11	Q9D9Q5	Q9d9q5 mus musculu
306	6	1.5	79	5	Q9TX84	Q9tx84 caenorhabdi	379	6	1.5	111	11	Q9CQ70	Q9cq70 mus musculu
307	6	1.5	79	6	Q9GURS	Q9gurs sus scrofa	380	6	1.5	112	2	Q9S6Q8	Q9s6q8 escherichia
308	6	1.5	79	6	Q8WNQ9	Q8wnq9 macaca mula	381	6	1.5	112	11	Q9D8T7	Q9d8t7 mus musculu

382	6	1.5	112	16	Q97SV7	Q97sv7 streptococ	455	6	1.5	127	5	002000	O02000 chrysaora q
383	6	1.5	112	17	Q970H6	Q970h6 sulfolobus	456	6	1.5	127	16	Q8RF26	Q8rf26 fusobacteri
384	6	1.5	115	15	Q9Q799	Q9q799 human immun	457	6	1.5	127	17	Q9YDY6	Q9ydy6 aeropyrum p
385	6	1.5	115	15	Q9Q798	Q9q798 human immun	458	6	1.5	127	17	Q96ZM5	Q96zm5 sulfolobus
386	6	1.5	115	15	Q9Q797	Q9q797 human immun	459	6	1.5	127	17	Q8TSM1	Q8tsw1 methanosaarc
387	6	1.5	115	15	Q9Q796	Q9q796 human immun	460	6	1.5	128	8	Q9GAB1	Q9gab1 calyptrogena
388	6	1.5	115	15	Q9Q795	Q9q795 human immun	461	6	1.5	129	4	Q9H8S0	Q9h8s0 homo sapien
389	6	1.5	115	15	Q9Q794	Q9q794 human immun	462	6	1.5	129	11	Q9ESG1	Q9esg1 mus musculus
390	6	1.5	115	15	Q9Q793	Q9q793 human immun	463	6	1.5	130	2	Q44320	Q44320 agrobacteri
391	6	1.5	115	15	Q9Q792	Q9q792 human immun	464	6	1.5	130	4	Q9H541	Q9h541 homo sapien
392	6	1.5	115	15	Q9Q791	Q9q791 human immun	465	6	1.5	130	5	Q8T3X9	Q8t3x9 drosophila
393	6	1.5	115	15	Q9Q790	Q9q790 human immun	466	6	1.5	130	8	Q8SG31	Q8sg31 nycteribid
394	6	1.5	115	15	Q9Q789	Q9q789 human immun	467	6	1.5	131	3	Q9U066	Q9u066 schizosacch
395	6	1.5	115	15	Q9Q788	Q9q788 human immun	468	6	1.5	131	4	Q9Y3Y0	Q9y3y0 homo sapien
396	6	1.5	115	15	Q9Q787	Q9q787 human immun	469	6	1.5	131	4	Q9NX35	Q9nx35 homo sapien
397	6	1.5	115	15	Q9Q786	Q9q786 human immun	470	6	1.5	131	5	Q9VXR0	Q9vxr0 drosophila
398	6	1.5	115	15	Q9Q785	Q9q785 human immun	471	6	1.5	131	16	Q8YXV4	Q8yxv4 anabaena sp
399	6	1.5	115	15	Q9Q784	Q9q784 human immun	472	6	1.5	134	2	Q93MW1	Q93mw1 nocardia br
400	6	1.5	115	15	Q9Q783	Q9q783 human immun	473	6	1.5	134	16	Q8XKP1	Q8xkp1 clostridium
401	6	1.5	115	15	Q9Q782	Q9q782 human immun	474	6	1.5	135	4	Q9NT69	Q9nt69 homo sapien
402	6	1.5	115	15	Q9Q781	Q9q781 human immun	475	6	1.5	135	13	Q8UVJ5	Q8uvj5 fugu rubrip
403	6	1.5	115	15	Q9Q780	Q9q780 human immun	476	6	1.5	135	16	P95252	P95252 mycobacteri
404	6	1.5	115	15	Q9Q779	Q9q779 human immun	477	6	1.5	136	5	Q9GNT2	Q9gnt2 stoichactis
405	6	1.5	115	15	Q9Q778	Q9q778 human immun	478	6	1.5	136	16	Q9EH08	Q9eh08 rhizobium l
406	6	1.5	115	15	Q9Q777	Q9q777 human immun	479	6	1.5	136	16	Q8YWI4	Q8ywi4 anabaena sp
407	6	1.5	115	15	Q9Q776	Q9q776 human immun	480	6	1.5	137	11	Q35946	Q35946 rattus norv
408	6	1.5	115	15	Q9Q775	Q9q775 human immun	481	6	1.5	137	16	Q8ZS75	Q8z875 anabaena sp
409	6	1.5	115	15	Q9Q774	Q9q774 human immun	482	6	1.5	137	16	Q8Z6A3	Q8z6a3 salmonella
410	6	1.5	115	15	Q9Q773	Q9q773 human immun	483	6	1.5	137	16	Q8RIS3	Q8ria3 streptomyc
411	6	1.5	115	15	Q9Q772	Q9q772 human immun	484	6	1.5	139	4	Q8TAU6	Q8taue homo sapien
412	6	1.5	115	15	Q9Q771	Q9q771 human immun	485	6	1.5	139	10	Q9AU91	Q9au91 catopsis mo
413	6	1.5	115	15	Q9Q770	Q9q770 human immun	486	6	1.5	140	10	Q9SUN2	Q9sun2 arabidopsis
414	6	1.5	115	15	Q9Q769	Q9q769 human immun	487	6	1.5	140	12	Q9EMS9	Q9em9 amacta moo
415	6	1.5	115	15	Q9Q768	Q9q768 human immun	488	6	1.5	140	16	Q8UEZ5	Q8uez5 agrobacteri
416	6	1.5	115	15	Q9Q767	Q9q767 human immun	489	6	1.5	140	16	Q8R7U2	Q8r7u2 thermoanaer
417	6	1.5	115	15	Q9Q766	Q9q766 human immun	490	6	1.5	141	2	Q59479	Q59479 klebsiella
418	6	1.5	115	15	Q9Q765	Q9q765 human immun	491	6	1.5	142	1	Q9HH43	Q9hh43 methanobact
419	6	1.5	115	15	Q9Q764	Q9q764 human immun	492	6	1.5	143	16	Q97DQ9	Q97dq9 clostridium
420	6	1.5	115	15	Q9Q763	Q9q763 human immun	493	6	1.5	144	4	Q9NT46	Q9nt46 homo sapien
421	6	1.5	117	2	Q9KKC9	Q9kkc9 borrelia he	494	6	1.5	144	10	Q9M5L1	Q9m5l1 euphorbia e
422	6	1.5	117	16	Q9KDB9	Q9kdb9 bacillus ha	495	6	1.5	144	16	Q99S28	Q99s28 staphylococ
423	6	1.5	117	16	Q8YSF9	Q8ysf9 anabaena sp	496	6	1.5	144	16	Q66910	Q66910 aquifex aeo
424	6	1.5	118	11	Q99J70	Q99j70 mus musculu	497	6	1.5	145	8	Q9TA91	Q9ta91 oeneis chry
425	6	1.5	118	12	Q99GR0	Q99gr0 culex nigri	498	6	1.5	145	8	Q9TA90	Q9ta90 oeneis lval
426	6	1.5	118	16	Q840A0	Q840a0 chlamydia c	499	6	1.5	145	8	Q9TA89	Q9ta89 oeneis chry
427	6	1.5	119	8	Q8SH40	Q8sh40 gasterophil	500	6	1.5	145	8	Q9G7Q0	Q9g7q0 teloglabrus
428	6	1.5	120	12	Q9J3A3	Q9j3a3 hepatitis c	501	6	1.5	145	8	Q9G7P4	Q9g7p4 diasmopsis
429	6	1.5	120	16	Q9PGS4	Q9pgs4 xylella fas	502	6	1.5	145	16	Q8UBQ4	Q8ubq4 agrobacteri
430	6	1.5	120	17	Q58480	Q58480 pyrococcus	503	6	1.5	146	8	Q9T3C0	Q9t3c0 oeneis chry
431	6	1.5	121	6	Q9BDP6	Q9bdp6 sus scrofa	504	6	1.5	146	8	Q9T3B9	Q9t3b9 oeneis lval
432	6	1.5	121	8	Q48202	Q48202 drosophila	505	6	1.5	146	8	Q9T3B8	Q9t3b8 oeneis chry
433	6	1.5	121	12	Q89229	Q89229 variola vir	506	6	1.5	146	8	Q9T312	Q9t312 oeneis neva
434	6	1.5	122	8	Q8SG41	Q8sg41 ectinorhync	507	6	1.5	146	8	Q9TEB7	Q9teb7 oeneis neva
435	6	1.5	122	8	Q8SG33	Q8sg33 gasterophil	508	6	1.5	146	8	Q9TEB6	Q9teb6 oeneis chry
436	6	1.5	122	10	Q94181	Q94181 oryza sativ	509	6	1.5	146	8	Q9TE65	Q9te65 oeneis sp.
437	6	1.5	122	13	Q90WW2	Q90ww2 xenopus lae	510	6	1.5	146	8	Q9TE64	Q9te64 oeneis sp.
438	6	1.5	122	16	Q9KSF7	Q9ksf7 vibrio chol	511	6	1.5	146	10	Q9MBX9	Q9mbx9 arabidopsis
439	6	1.5	122	17	Q97B17	Q97bi7 thermoplas	512	6	1.5	146	10	Q9SK22	Q9sk22 arabidopsis
440	6	1.5	123	4	Q96RL5	Q96rl5 homo sapien	513	6	1.5	146	10	Q9S994	Q9s994 phaseolus v
441	6	1.5	123	8	Q9VY37	Q9vy37 drosophila	514	6	1.5	146	17	Q9HT63	Q9ht63 thermoplas
442	6	1.5	123	8	Q9KH57	Q9kh57 pinus sp. q	515	6	1.5	147	2	Q9EUM5	Q9eum5 corynebacte
443	6	1.5	123	8	Q8SH46	Q8sh46 ectinorhync	516	6	1.5	147	6	Q27996	Q27996 bos taurus
444	6	1.5	123	10	Q42253	Q42253 arabidopsis	517	6	1.5	147	8	Q9MJ41	Q9mj41 rana nigrom
445	6	1.5	123	16	Q9A2E4	Q9a2e4 caulobacter	518	6	1.5	147	8	Q9MJ40	Q9mj40 rana porosa
446	6	1.5	124	2	Q45924	Q45924 clostridium	519	6	1.5	147	8	Q9MJ39	Q9mj39 rana porosa
447	6	1.5	124	8	Q03898	Q03898 enchenopa b	520	6	1.5	147	8	Q9MJ38	Q9mj38 rana porosa
448	6	1.5	124	16	Q9X8X6	Q9x8x6 streptomyc	521	6	1.5	147	8	Q9MJ37	Q9mj37 rana porosa
449	6	1.5	124	16	Q8R877	Q8r8j7 thermoanaer	522	6	1.5	147	10	Q9LKU2	Q9lkuz arabidopsis
450	6	1.5	124	17	Q59587	Q59587 pyrococcus	523	6	1.5	147	16	Q8XQH5	Q8xqh5 ralatonia s
451	6	1.5	125	12	Q9J3B0	Q9j3b0 hepatitis c	524	6	1.5	147	16	Q9Z8L3	Q9z8l3 chlamydia p
452	6	1.5	126	2	Q9F0K6	Q9f0k6 staphylococ	525	6	1.5	148	2	Q93QR1	Q93qr1 elwinia pai
453	6	1.5	126	2	Q48876	Q48876 leptospira	526	6	1.5	148	2	Q8RJ32	Q8rj32 uncultured
454	6	1.5	126	16	Q99SM6	Q99sm6 staphylococ	527	6	1.5	148	5	Q9XUG4	Q9xug4 caenorhabdi

528	6	1.5	148	5	Q95W23	Q95W23 anthomomus	601	6	1.5	170	8	Q95990	Q95990 tegeticula
529	6	1.5	148	16	Q8XW23	Q8XW23 ralsstoma s	602	6	1.5	170	8	Q34614	Q34614 greya obacu
530	6	1.5	150	5	Q9VTLH	Q9VTLH drosophila	603	6	1.5	171	16	Q8ZRA3	Q8ZRA3 salmoneila
531	6	1.5	150	16	Q34682	Q34682 bacillus su	604	6	1.5	172	8	Q95984	Q95984 tegeticula
532	6	1.5	150	17	Q9YDY3	Q9YDY3 aeropyrum p	605	6	1.5	172	16	Q8ZAT2	Q8ZAT2 yersinia pe
533	6	1.5	150	17	Q8ZSW2	Q8ZSW2 pyrobaculum	606	6	1.5	172	16	Q8XV29	Q8XV29 ralsstoma s
534	6	1.5	152	5	Q26590	Q26590 tritrichomo	607	6	1.5	172	17	Q8U4C2	Q8U4C2 pyrococcus
535	6	1.5	152	11	Q92013	Q92013 mus musculu	608	6	1.5	173	10	Q9CAW4	Q9CAW4 arabiidopsis
536	6	1.5	152	17	Q30299	Q30299 archaeglob	609	6	1.5	174	8	Q9G0Y7	Q9G0Y7 tegeticula
537	6	1.5	153	16	Q8XCZ1	Q8XCZ1 escherichia	610	6	1.5	174	8	Q9TC03	Q9TC03 perthida gl
538	6	1.5	154	2	Q939S0	Q939S0 dehalospiri	611	6	1.5	174	8	Q9TBZ9	Q9TBZ9 greya solen
539	6	1.5	154	12	Q916X1	Q916X1 hepatitis c	612	6	1.5	174	8	Q9TBZ7	Q9TBZ7 greya punct
540	6	1.5	154	12	Q916W9	Q916W9 hepatitis c	613	6	1.5	174	8	Q9TBZ5	Q9TBZ5 lampironia a
541	6	1.5	154	12	Q916W7	Q916W7 hepatitis c	614	6	1.5	174	8	Q9TBZ1	Q9TBZ1 mesepiola s
542	6	1.5	154	12	Q916W6	Q916W6 hepatitis c	615	6	1.5	174	8	Q9TBZ9	Q9TBZ9 prodocus ae
543	6	1.5	154	12	Q8QON2	Q8QON2 hepatitis c	616	6	1.5	174	8	Q9TBV7	Q9TBV7 prodocus ma
544	6	1.5	154	12	Q8QON2	Q8QON2 hepatitis c	617	6	1.5	174	8	Q9TBV5	Q9TBV5 prodocus co
545	6	1.5	154	16	Q8RF84	Q8RF84 fusobacteri	618	6	1.5	174	8	Q9TBV3	Q9TBV3 prodocus y-
546	6	1.5	155	13	Q91459	Q91459 seriola qui	619	6	1.5	174	8	Q9TBV1	Q9TBV1 prodocus pa
547	6	1.5	155	16	Q9HY74	Q9HY74 pseudomonas	620	6	1.5	174	8	Q9TBX9	Q9TBX9 prodocus sp
548	6	1.5	155	5	Q9BPM2	Q9BPM2 metagonimus	621	6	1.5	174	8	Q9TBX5	Q9TBX5 parategetic
549	6	1.5	156	16	Q926Q2	Q926Q2 listeria in	622	6	1.5	174	8	Q9TBX1	Q9TBX1 yersinia que
550	6	1.5	156	16	Q8Y3T0	Q8Y3T0 listeria mo	623	6	1.5	174	8	Q9TBW9	Q9TBW9 incurvata
551	6	1.5	156	16	Q8XLU8	Q8XLU8 clostridium	624	6	1.5	174	8	Q9TBW7	Q9TBW7 greya polit
552	6	1.5	156	16	Q8UTU1	Q8UTU1 agrobacteri	625	6	1.5	174	8	Q95810	Q95810 greya polit
553	6	1.5	156	17	Q9Y9D6	Q9Y9D6 aeropyrum p	626	6	1.5	174	8	Q95959	Q95959 tegeticula
554	6	1.5	157	8	Q9T504	Q9T504 lucilia cup	627	6	1.5	174	8	Q95961	Q95961 tegeticula
555	6	1.5	157	11	Q9ZSU6	Q9ZSU6 oryza sativ	628	6	1.5	174	8	Q9G7B2	Q9G7B2 tegeticula
556	6	1.5	157	11	Q8R4H5	Q8R4H5 rattus norv	629	6	1.5	174	8	Q9G7B0	Q9G7B0 tegeticula
557	6	1.5	157	16	Q8RRJ0	Q8RRJ0 delnoccocus	630	6	1.5	174	8	Q9G7B0	Q9G7B0 tegeticula
558	6	1.5	157	16	Q9YBM9	Q9YBM9 aeropyrum p	631	6	1.5	174	8	Q9G3L5	Q9G3L5 tegeticula
559	6	1.5	158	9	Q9B067	Q9B067 mycobacteri	632	6	1.5	174	8	Q9G3L3	Q9G3L3 tegeticula
560	6	1.5	158	10	Q9AXN8	Q9AXN8 physcomitre	633	6	1.5	174	8	Q9G3K9	Q9G3K9 tegeticula
561	6	1.5	159	2	Q9EUI4	Q9EUI4 salmoneila	634	6	1.5	174	8	Q9G3K7	Q9G3K7 tegeticula
562	6	1.5	159	4	Q9NXY8	Q9NXY8 homo sapien	635	6	1.5	174	8	Q9G3K5	Q9G3K5 tegeticula
563	6	1.5	159	8	P92539	P92539 arabiidopsis	636	6	1.5	174	8	Q9G3K3	Q9G3K3 tegeticula
564	6	1.5	159	10	Q8W0T0	Q8W0T0 scorghum bic	637	6	1.5	174	8	Q9G3K1	Q9G3K1 tegeticula
565	6	1.5	159	10	Q85683	Q85683 arabiidopsis	638	6	1.5	174	8	Q9G7A8	Q9G7A8 tegeticula
566	6	1.5	159	10	Q858C5	Q858C5 arabiidopsis	639	6	1.5	174	8	Q9G7A6	Q9G7A6 tegeticula
567	6	1.5	159	10	Q85876	Q85876 arabiidopsis	640	6	1.5	174	8	Q9G7A4	Q9G7A4 prodocus n.
568	6	1.5	159	16	Q8Y8B2	Q8Y8B2 listeria mo	641	6	1.5	174	8	Q9G3J9	Q9G3J9 tegeticula
569	6	1.5	159	17	Q9Y9E7	Q9Y9E7 aeropyrum p	642	6	1.5	174	8	Q95994	Q95994 tegeticula
570	6	1.5	159	17	Q8ZXR8	Q8ZXR8 pyrobaculum	643	6	1.5	174	8	Q95978	Q95978 tegeticula
571	6	1.5	159	17	Q8ZXR8	Q8ZXR8 methanosarc	644	6	1.5	174	8	Q95982	Q95982 tegeticula
572	6	1.5	159	17	Q8ZSP9	Q8ZSP9 methanosarc	645	6	1.5	174	8	Q95986	Q95986 tegeticula
573	6	1.5	160	10	Q43481	Q43481 bordum vul	646	6	1.5	174	8	Q95988	Q95988 tegeticula
574	6	1.5	160	10	P94060	P94060 arabiidopsis	647	6	1.5	174	8	Q95992	Q95992 tegeticula
575	6	1.5	160	17	Q9HM01	Q9HM01 thermoplasm	648	6	1.5	174	8	Q95965	Q95965 tegeticula
576	6	1.5	161	5	Q9N9S6	Q9N9S6 leishmania	649	6	1.5	174	8	Q95968	Q95968 tegeticula
577	6	1.5	161	10	Q81095	Q81095 leaquereilla	650	6	1.5	174	8	Q95973	Q95973 tegeticula
578	6	1.5	161	15	Q9U026	Q9U026 human immun	651	6	1.5	174	8	Q95976	Q95976 tegeticula
579	6	1.5	161	16	Q9RSZ0	Q9RSZ0 delnoccocus	652	6	1.5	174	8	Q95976	Q95976 tegeticula
580	6	1.5	161	16	Q9RSZ0	Q9RSZ0 staphylococ	653	6	1.5	174	8	Q95986	Q95986 tegeticula
581	6	1.5	161	17	Q8TUX7	Q8TUX7 methanopyru	654	6	1.5	174	8	Q95996	Q95996 tegeticula
582	6	1.5	162	5	Q02256	Q02256 caenorhabdi	655	6	1.5	174	10	Q9ATW5	Q9ATW5 zea mays (m
583	6	1.5	162	5	Q95X62	Q95X62 caenorhabdi	656	6	1.5	174	16	Q9ATW0	Q9ATW0 bacillus ha
584	6	1.5	162	5	Q8T100	Q8T100 bombyx mori	657	6	1.5	175	8	Q9XW02	Q9XW02 tegeticula
585	6	1.5	162	16	Q9ZH00	Q9ZH00 rickettsia	658	6	1.5	175	8	Q9XW02	Q9XW02 tegeticula
586	6	1.5	163	2	Q93RF3	Q93RF3 treponema m	659	6	1.5	175	16	Q9PEN2	Q9PEN2 xylella las
587	6	1.5	163	2	Q923W3	Q923W3 pseudomonas	660	6	1.5	175	16	Q9CDU0	Q9CDU0 lactococcus
588	6	1.5	163	5	Q9BMR6	Q9BMR6 brugia mala	661	6	1.5	176	2	Q93Q41	Q93Q41 myxococcus
589	6	1.5	163	5	Q9BHV2	Q9BHV2 leishmania	662	6	1.5	176	4	Q9H8F6	Q9H8F6 homo sapien
590	6	1.5	164	5	Q9VLE4	Q9VLE4 drosophila	663	6	1.5	176	4	Q9H8A6	Q9H8A6 homo sapien
591	6	1.5	164	16	Q9S2D0	Q9S2D0 streptomyce	664	6	1.5	176	8	Q34619	Q34619 greya obacu
592	6	1.5	166	2	Q8VM87	Q8VM87 rhizobium l	665	6	1.5	176	11	Q9DC76	Q9DC76 mus musculu
593	6	1.5	166	5	Q99281	Q99281 parechilinus	666	6	1.5	176	16	Q9A6B0	Q9A6B0 caulobacter
594	6	1.5	167	16	Q92237	Q92237 rhizobium m	667	6	1.5	176	16	Q930V6	Q930V6 rhizobium m
595	6	1.5	168	16	Q8R6B0	Q8R6B0 fusobacteri	668	6	1.5	176	17	Q9HK81	Q9HK81 thermoplasm
596	6	1.5	169	16	Q8X8U0	Q8X8U0 escherichia	669	6	1.5	177	4	Q9HBD1	Q9HBD1 homo sapien
597	6	1.5	169	17	Q9Y976	Q9Y976 aeropyrum p	670	6	1.5	178	9	Q93040	Q93040 tridentator
598	6	1.5	169	17	Q9HSA7	Q9HSA7 halobacteri	671	6	1.5	178	9	Q8SB06	Q8SB06 bacteriopho
599	6	1.5	170	2	Q9K539	Q9K539 mycobacteri	672	6	1.5	178	10	Q49595	Q49595 arabiidopsis
600	6	1.5	170	8	Q95980	Q95980 tegeticula	673	6	1.5	179	8	Q47368	Q47368 gonioctena

674	6	1.5	179	8	Q9GAH1	Q9GAH1 melophorus	747	6	1.5	188	12	Q9J3A4	Q9J3A4 hepatitis c
675	6	1.5	180	5	O76362	O76362 caenorhabdi	748	6	1.5	188	12	Q9J3A5	Q9J3A5 hepatitis c
676	6	1.5	180	8	O47374	O47374 oreina caca	749	6	1.5	188	12	Q9J3A6	Q9J3A6 hepatitis c
677	6	1.5	180	8	O47357	O47357 gonioctena	750	6	1.5	188	12	Q9J3A8	Q9J3A8 hepatitis c
678	6	1.5	180	8	O47358	O47358 gonioctena	751	6	1.5	188	12	Q9J3B1	Q9J3B1 hepatitis c
679	6	1.5	180	8	O47359	O47359 gonioctena	752	6	1.5	188	12	Q9J3B2	Q9J3B2 hepatitis c
680	6	1.5	180	8	O47360	O47360 gonioctena	753	6	1.5	188	12	Q9J3B3	Q9J3B3 hepatitis c
681	6	1.5	180	8	O47361	O47361 gonioctena	754	6	1.5	188	12	Q9J3B4	Q9J3B4 hepatitis c
682	6	1.5	180	8	O47362	O47362 gonioctena	755	6	1.5	188	12	Q9J3B5	Q9J3B5 hepatitis c
683	6	1.5	180	8	O47363	O47363 gonioctena	756	6	1.5	189	4	Q8MU39	Q8MU39 homo sapien
684	6	1.5	180	8	O47364	O47364 gonioctena	757	6	1.5	189	8	Q9MF30	Q9MF30 brachydes
685	6	1.5	180	8	O47365	O47365 gonioctena	758	6	1.5	189	8	Q9MF32	Q9MF32 brachydes
686	6	1.5	180	8	O47366	O47366 gonioctena	759	6	1.5	189	10	Q9SKV7	Q9SKV7 lithospermu
687	6	1.5	180	8	O47367	O47367 gonioctena	760	6	1.5	189	10	Q9MA39	Q9MA39 arabidopsis
688	6	1.5	180	8	O47369	O47369 gonioctena	761	6	1.5	190	8	Q9MDU5	Q9MDU5 brachydes
689	6	1.5	180	8	O47370	O47370 gonioctena	762	6	1.5	190	8	Q9MDT3	Q9MDT3 brachydes
690	6	1.5	180	8	O47371	O47371 gonioctena	763	6	1.5	190	8	Q9ME27	Q9ME27 brachydes
691	6	1.5	180	8	O47372	O47372 gonioctena	764	6	1.5	190	8	Q9MF02	Q9MF02 brachydes
692	6	1.5	180	8	O47373	O47373 gonioctena	765	6	1.5	190	8	Q9MF16	Q9MF16 brachydes
693	6	1.5	180	8	O47375	O47375 chrysomela	766	6	1.5	190	8	Q9MF29	Q9MF29 brachydes
694	6	1.5	180	8	O47362	Q34562 greya solen	767	6	1.5	190	8	Q9MF26	Q9MF26 brachydes
695	6	1.5	180	8	O47356	Q34564 greya enchr	768	6	1.5	190	8	Q9MF21	Q9MF21 brachydes
696	6	1.5	180	8	O47359	Q34594 greya mitel	769	6	1.5	190	8	Q9MF12	Q9MF12 brachydes
697	6	1.5	180	8	Q34617	Q34617 greya obscu	770	6	1.5	190	10	Q94J85	Q94J85 cryza activ
698	6	1.5	180	8	Q34621	Q34621 greya obscu	771	6	1.5	190	16	Q8ZJV6	Q8ZJV6 salmonella
699	6	1.5	180	8	Q34627	Q34627 greya polit	772	6	1.5	191	5	Q9U8C2	Q9U8C2 leishmania
700	6	1.5	180	8	Q34629	Q34629 greya powel	773	6	1.5	191	16	Q8Z123	Q8Z123 yersinia pe
701	6	1.5	180	8	Q34631	Q34631 greya polit	774	6	1.5	192	8	Q9B246	Q9B246 chrysomela
702	6	1.5	180	8	Q34633	Q34633 greya pecti	775	6	1.5	192	8	Q94XM3	Q94XM3 haliphthoro
703	6	1.5	180	8	Q34639	Q34639 greya piper	776	6	1.5	192	8	Q94XM1	Q94XM1 haliphthoro
704	6	1.5	180	8	Q34655	Q34655 greya retic	777	6	1.5	192	8	Q8WEQ8	Q8WEQ8 lepenis har
705	6	1.5	180	8	Q34663	Q34663 greya subal	778	6	1.5	193	3	Q12064	Q12064 saccharomyc
706	6	1.5	180	8	Q34665	Q34665 greya subal	779	6	1.5	193	5	Q9GUB2	Q9GUB2 litomosolide
707	6	1.5	180	8	Q34667	Q34667 greya subal	780	6	1.5	193	8	Q9MIU5	Q9MIU5 sapromyces
708	6	1.5	180	8	Q34679	Q34679 greya varia	781	6	1.5	193	8	Q9B256	Q9B256 chrysomela
709	6	1.5	180	8	Q36678	Q36678 prodoxus qu	782	6	1.5	193	8	Q94RN1	Q94RN1 platyphora
710	6	1.5	180	8	Q34641	Q34641 greya piper	783	6	1.5	193	8	Q9MIU6	Q9MIU6 pythium ult
711	6	1.5	180	8	Q48288	Q48288 gonioctena	784	6	1.5	194	4	Q96FL4	Q96FL4 homo sapien
712	6	1.5	180	10	O65505	O65505 arabidopsis	785	6	1.5	194	5	O76944	O76944 fasciola he
713	6	1.5	180	16	Q91085	Q91085 pseudomonas	786	6	1.5	194	8	Q9B258	Q9B258 chrysomela
714	6	1.5	181	5	Q9NH98	Q9NH98 stylonychia	787	6	1.5	194	8	Q9B250	Q9B250 chrysomela
715	6	1.5	181	10	Q9XGT3	Q9XGT3 vitis vinif	788	6	1.5	194	8	Q95918	Q95918 gnorimoph
716	6	1.5	181	10	Q9XGT2	Q9XGT2 arabidopsis	789	6	1.5	194	16	P73101	P73101 synchocyst
717	6	1.5	182	3	Q33520	Q33520 saccharomyc	790	6	1.5	194	17	Q9YB78	Q9YB78 aeropyrum p
718	6	1.5	182	4	Q9NQP9	Q9NQP9 homo sapien	791	6	1.5	195	10	Q9FHC1	Q9FHC1 arabidopsis
719	6	1.5	182	9	Q9JMN5	Q9JMN5 bacterioph	792	6	1.5	195	16	Q97DF9	Q97DF9 clostridium
720	6	1.5	184	8	Q9C163	Q9C163 dendroctonu	793	6	1.5	195	16	Q97DF9	Q97DF9 homo sapien
721	6	1.5	184	8	Q9G828	Q9G828 dendroctonu	794	6	1.5	196	4	Q96CQ2	Q96CQ2 homo sapien
722	6	1.5	184	8	Q9C159	Q9C159 greya powel	795	6	1.5	196	8	Q9B254	Q9B254 chrysomela
723	6	1.5	184	8	Q9C161	Q9C161 greya powel	796	6	1.5	196	8	Q94RN8	Q94RN8 platyphora
724	6	1.5	184	8	Q9C163	Q9C163 greya solen	797	6	1.5	196	8	Q94RM3	Q94RM3 platyphora
725	6	1.5	184	8	Q9C164	Q9C164 greya solen	798	6	1.5	196	8	Q9T226	Q9T226 phytophthor
726	6	1.5	184	16	Q99Y12	Q99Y12 atreptococc	799	6	1.5	196	11	O88887	O88887 mus musculu
727	6	1.5	185	8	Q33831	Q33831 prodoxus pa	800	6	1.5	197	8	Q9B261	Q9B261 chrysomela
728	6	1.5	185	8	Q34855	Q34855 lampronia a	801	6	1.5	197	8	Q9B251	Q9B251 chrysomela
729	6	1.5	185	8	Q35352	Q35352 prodoxus ae	802	6	1.5	197	8	Q9B245	Q9B245 chrysomela
730	6	1.5	185	8	Q35531	Q35531 prodoxus ma	803	6	1.5	197	8	Q94RP5	Q94RP5 platyphora
731	6	1.5	185	8	Q35601	Q35601 parategetic	804	6	1.5	197	8	Q94RN9	Q94RN9 platyphora
732	6	1.5	185	8	Q36080	Q36080 tegeticula	805	6	1.5	197	8	Q94RN6	Q94RN6 stilodora fu
733	6	1.5	185	8	Q35095	Q35095 meseptiola s	806	6	1.5	197	8	Q94RM9	Q94RM9 platyphora
734	6	1.5	185	8	Q36184	Q36184 tegeticula	807	6	1.5	197	11	Q9ERC3	Q9ERC3 mus musculu
735	6	1.5	185	16	Q8X704	Q8X704 escherichia	808	6	1.5	197	15	Q9EBM2	Q9EBM2 human immun
736	6	1.5	185	16	Q8U120	Q8U120 rhodobacteri	809	6	1.5	197	16	Q8Y777	Q8Y777 anabaena ep
737	6	1.5	186	2	Q9R17	Q9R17 rhodobacteri	810	6	1.5	197	16	Q8Y777	Q8Y777 anabaena ep
738	6	1.5	186	3	Q9C3Y6	Q9C3Y6 neurospora	811	6	1.5	198	5	Q9W1R8	Q9W1R8 drosophila
739	6	1.5	186	3	Q9CTV7	Q9CTV7 drosophila	812	6	1.5	198	5	Q8T924	Q8T924 drosophila
740	6	1.5	186	10	Q9ZRF4	Q9ZRF4 nicotiana c	813	6	1.5	198	8	Q9B1J7	Q9B1J7 chrysomela
741	6	1.5	186	10	Q9SKR9	Q9SKR9 arabidopsis	814	6	1.5	198	15	Q9EBM0	Q9EBM0 human immun
742	6	1.5	187	3	Q9US85	Q9US85 schizosacch	815	6	1.5	198	16	Q9R1W5	Q9R1W5 streptomyces
743	6	1.5	187	5	Q17580	Q17580 caenorhabdi	816	6	1.5	199	4	Q8WYU9	Q8WYU9 homo sapien
744	6	1.5	187	16	Q8ZMP8	Q8ZMP8 salmonella	817	6	1.5	199	5	Q9GUH8	Q9GUH8 caenorhabdi
745	6	1.5	187	16	Q8X503	Q8X503 escherichia	818	6	1.5	199	5	O16006	O16006 dirofilaria
746	6	1.5	188	11	Q9Z2E6	Q9Z2E6 mus musculu	819	6	1.5	199	5	O16026	O16026 dirofilaria

820	6	1.5	199	8	Q94RN4	Q94tn4 platyphora	893	6	1.5	208	12	Q9YUF8	Q9Yuf8 rhesus cyto
821	6	1.5	199	10	Q9F174	Q9f1j4 arabidopsis	894	6	1.5	208	13	Q9PSA1	Q9psa1 gallus gall
822	6	1.5	199	11	Q9CQJ7	Q9cjq7 mus musculus	895	6	1.5	208	16	Q9SESO	Q9seso rhizobium 1
823	6	1.5	199	11	P97613	P97613 rattus norv	896	6	1.5	209	4	Q96SH7	Q96sh7 homo sapien
824	6	1.5	199	15	Q9EBL9	Q9eb19 human immun	897	6	1.5	209	10	Q9ZNX3	Q9znx3 pecunia hyb
825	6	1.5	200	6	Q9N1N8	Q9n1n8 musceta put	898	6	1.5	209	11	Q9CXU4	Q9cxu4 mus musculus
826	6	1.5	200	8	Q9B247	Q9b247 chrysomela	899	6	1.5	209	15	P90041	P90041 human immun
827	6	1.5	200	8	Q94RN5	Q94rn5 stilloides mo	900	6	1.5	209	15	Q96924	Q96924 human immun
828	6	1.5	200	8	Q94RM6	Q94rm6 platyphora	901	6	1.5	209	15	Q9XA31	Q9xa31 streptomyc
829	6	1.5	201	8	Q94RP2	Q94rp2 platyphora	902	6	1.5	211	5	Q9BLE4	Q9ble4 entamoeba h
830	6	1.5	201	8	Q94RP0	Q94rp0 platyphora	903	6	1.5	211	5	Q18532	Q18532 onchocerca
831	6	1.5	201	8	Q94RN7	Q94rn7 calligrapha	904	6	1.5	211	8	Q9G899	Q9g899 speyeria id
832	6	1.5	201	8	Q94RM8	Q94rm8 platyphora	905	6	1.5	211	11	Q9ESH5	Q9esh5 mus musculus
833	6	1.5	201	8	Q94RM4	Q94rm4 platyphora	906	6	1.5	212	5	Q9N712	Q9n712 leishmania
834	6	1.5	201	11	Q9CQZ8	Q9cqz8 mus musculus	907	6	1.5	212	6	Q9SL41	Q9sl41 equus caball
835	6	1.5	201	15	Q8OC70	Q8oct0 human immun	908	6	1.5	212	11	Q70280	Q70280 rattus norv
836	6	1.5	202	8	Q9B253	Q9b253 chrysomela	909	6	1.5	212	16	Q9RT55	Q9rt55 deinococcus
837	6	1.5	202	8	Q9B248	Q9b248 chrysomela	910	6	1.5	212	16	Q53457	Q53457 mycobacteri
838	6	1.5	202	8	Q94PQ7	Q94pq7 anartia iyt	911	6	1.5	213	6	Q9SL42	Q9sl42 sus scrofa
839	6	1.5	202	8	Q94PC0	Q94pc0 anartia ama	912	6	1.5	213	16	Q8Y0Q9	Q8y0q9 ralsconia s
840	6	1.5	202	8	Q94N05	Q94nq5 anartia chr	913	6	1.5	214	8	Q01925	Q01925 glycine max
841	6	1.5	202	8	Q94NFO	Q94nfo anartia jat	914	6	1.5	214	16	Q8YVU3	Q8yvut3 anabaena sp
842	6	1.5	202	15	Q80697	Q80697 human immun	915	6	1.5	215	2	P94879	P94879 lactococcus
843	6	1.5	202	16	Q9A473	Q9a473 callobacter	916	6	1.5	215	16	Q8XKC6	Q8xkc6 clostridium
844	6	1.5	202	16	Q9L287	Q9l287 streptomyc	917	6	1.5	215	17	Q8U013	Q8u013 pyrococcus
845	6	1.5	202	16	Q8Y1B5	Q8y1b5 bruceella me	918	6	1.5	216	8	Q37451	Q37451 dryas iulia
846	6	1.5	203	3	Q07451	Q07451 saccharomyc	919	6	1.5	216	8	Q37481	Q37481 heliconius
847	6	1.5	203	4	Q9UDV0	Q9udv0 homo sapien	920	6	1.5	216	10	Q9SM75	Q9sm75 cyanophora
848	6	1.5	203	5	Q9NCT3	Q9nct3 caenorhabdi	921	6	1.5	216	17	Q9UZD9	Q9uzd9 pyrococcus
849	6	1.5	203	5	Q9VRCU3	Q9vrcu3 drosophila	922	6	1.5	217	2	Q9LAF8	Q9laf8 bacillus th
850	6	1.5	203	5	Q8SMQ5	Q8smq5 enccephalito	923	6	1.5	217	4	Q9H7Q9	Q9h7q9 homo sapien
851	6	1.5	203	8	Q9B257	Q9b257 chrysomela	924	6	1.5	217	8	Q37530	Q37530 heliconius
852	6	1.5	203	8	Q9B252	Q9b252 chrysomela	925	6	1.5	217	16	Q67705	Q67705 aquifex aeo
853	6	1.5	203	8	Q9B249	Q9b249 chrysomela	926	6	1.5	218	16	Q8YMF5	Q8ymf5 anabaena sp
854	6	1.5	203	8	Q94RP1	Q94rp1 platyphora	927	6	1.5	218	8	Q37483	Q37483 heliconius
855	6	1.5	203	10	Q9FU77	Q9fuf7 oryza sativ	928	6	1.5	219	8	Q9B8G1	Q9b8g1 heterodoxus
856	6	1.5	204	8	Q63469	Q63469 pieris rapa	929	6	1.5	219	8	Q37502	Q37502 heliconius
857	6	1.5	204	8	Q94RM7	Q94rm7 platyphora	930	6	1.5	219	8	Q37493	Q37493 heliconius
858	6	1.5	204	10	Q9ASD1	Q9asd1 oryza sativ	931	6	1.5	219	8	Q63463	Q63463 cercyonia p
859	6	1.5	204	13	Q90YK4	Q9oyk4 sparus aura	932	6	1.5	219	16	Q9NWR0	Q9nwr0 straphylococ
860	6	1.5	204	13	Q90Y60	Q9oy60 sparus aura	933	6	1.5	220	3	Q42911	Q42911 schizosacch
861	6	1.5	204	13	Q90VV6	Q90vv6 epinephelus	934	6	1.5	220	4	Q9UG60	Q9ug60 homo sapien
862	6	1.5	204	13	Q8UW02	Q8uw02 epinephelus	935	6	1.5	220	5	Q9U2G1	Q9u2g1 caenorhabdi
863	6	1.5	204	16	Q8XFN7	Q8xfn7 salmoneella	936	6	1.5	220	8	Q37519	Q37519 heliconius
864	6	1.5	205	8	Q94XTU9	Q94xtu9 synemon pla	937	6	1.5	220	8	Q37488	Q37488 heliconius
865	6	1.5	205	8	Q94XTU8	Q94xtu8 synemon pla	938	6	1.5	220	8	Q37450	Q37450 dryas iulia
866	6	1.5	205	8	Q94RM5	Q94rm5 platyphora	939	6	1.5	220	8	Q37336	Q37336 heliconius
867	6	1.5	205	8	Q94NP9	Q94np9 synemon pla	940	6	1.5	220	8	Q63451	Q63451 actinote ge
868	6	1.5	205	8	Q94NP8	Q94np8 synemon pla	941	6	1.5	220	17	Q9HR64	Q9hr64 halobacteri
869	6	1.5	205	11	Q9CWN1	Q9cwn1 mus musculu	942	6	1.5	220	17	Q8ZTX2	Q8ztx2 pyrobaculum
870	6	1.5	205	16	Q9X782	Q9x782 mycobacteri	943	6	1.5	221	8	Q37531	Q37531 heliconius
871	6	1.5	206	5	Q15785	Q15785 entamoeba h	944	6	1.5	221	8	Q37337	Q37337 heliconius
872	6	1.5	206	11	Q9J1W9	Q9j1w9 mus musculu	945	6	1.5	221	8	Q37532	Q37532 heliconius
873	6	1.5	206	15	Q96922	Q96922 human immun	946	6	1.5	221	16	Q9JUL2	Q9jul2 neisseria m
874	6	1.5	206	16	Q9EWM6	Q9eww6 streptomyc	947	6	1.5	221	16	Q9HUI2	Q9hui2 pseudomonas
875	6	1.5	207	2	Q9RLM5	Q9rlm5 klebsiella	948	6	1.5	221	17	Q8UI86	Q8ui86 pyrococcus
876	6	1.5	207	2	Q56169	Q56169 streptomyc	949	6	1.5	222	5	Q95XC3	Q95xc3 caenorhabdi
877	6	1.5	207	4	Q9NKH4	Q9nkh4 homo sapien	950	6	1.5	222	5	Q9W418	Q9w418 drosophila
878	6	1.5	207	4	Q9GAK2	Q9gak2 kaloterme	951	6	1.5	222	8	Q9WGC7	Q9wgc7 panesthia c
879	6	1.5	207	8	Q9B260	Q9b260 chrysomela	952	6	1.5	222	8	Q9WGC4	Q9wgc4 serritermes
880	6	1.5	207	11	Q8RI10	Q8ri10 mus musculu	953	6	1.5	222	8	Q9WGC3	Q9wgc3 microhodote
881	6	1.5	207	16	Q9ABS1	Q9abs1 caulobacter	954	6	1.5	222	8	Q9WGC2	Q9wgc2 coplotermes
882	6	1.5	207	16	Q8Z163	Q8z163 yersinia pe	955	6	1.5	222	8	Q9B812	Q9b812 reticuliter
883	6	1.5	208	2	Q9APV0	Q9apv0 pseudomonas	956	6	1.5	222	8	Q94XT3	Q94xt3 psammoterm
884	6	1.5	208	2	Q48622	Q48622 lactococcus	957	6	1.5	222	8	Q94XT2	Q94xt2 prothioter
885	6	1.5	208	2	Q9ROB3	Q9rgb3 lactococcus	958	6	1.5	222	8	Q94XT1	Q94xt1 rhinoterme
886	6	1.5	208	8	Q9GAL6	Q9gal6 cryptococme	959	6	1.5	222	8	Q94XT0	Q94xt0 dolichorhin
887	6	1.5	208	8	Q9GAK7	Q9gak7 glyptocorme	960	6	1.5	222	8	Q94XS9	Q94xs9 schedorhino
888	6	1.5	208	8	Q9GAK1	Q9gak1 kaloterme	961	6	1.5	222	8	Q94XS8	Q94xs8 schedorhino
889	6	1.5	208	8	Q94RP6	Q94rp6 platyphora	962	6	1.5	222	8	Q94XS7	Q94xs7 parthioter
890	6	1.5	208	8	Q94RP3	Q94rp3 platyphora	963	6	1.5	222	8	Q94XS6	Q94xs6 parthioter
891	6	1.5	208	8	Q94RN3	Q94rn3 platyphora	964	6	1.5	222	8	Q94XS5	Q94xs5 reticuliter
892	6	1.5	208	8	Q94RN0	Q94rn0 platyphora	965	6	1.5	222	8	Q94XS4	Q94xs4 coplotermes

```

966 6 1.5 222 8 Q94XS3 094xs3 heteroterm
967 6 1.5 222 8 Q37755 037755 heliconius
968 6 1.5 222 8 Q37510 037510 heliconius
969 6 1.5 222 8 Q37512 037512 heliconius
970 6 1.5 222 8 Q37507 037507 heliconius
971 6 1.5 222 8 Q37492 037492 heliconius
972 6 1.5 222 8 Q37462 037462 eueides isa
973 6 1.5 222 8 Q37466 037466 eueides pro
974 6 1.5 222 8 Q37338 037338 heliconius
975 6 1.5 222 17 Q97XU3 097xu3 sulfolobus
976 6 1.5 223 2 Q49704 049704 mycobacteri
977 6 1.5 223 2 Q9AH22 09ah22 photorhabdu
978 6 1.5 223 2 Q8TZ6 08tz6 rhodopseudo
979 6 1.5 223 8 Q9MQ21 09mq21 laqueus rub
980 6 1.5 223 8 Q37752 037752 heliconius
981 6 1.5 223 8 Q37772 037772 heliconius
982 6 1.5 223 8 Q37814 037814 heliconius
983 6 1.5 223 8 Q37517 037517 heliconius
984 6 1.5 223 8 Q37508 037508 heliconius
985 6 1.5 223 8 Q37489 037489 heliconius
986 6 1.5 223 8 Q37558 037558 neruda aeod
987 6 1.5 223 8 Q63461 063461 morpho hele
988 6 1.5 223 8 Q03884 003884 bironella h
989 6 1.5 223 8 Q8W771 08w771 timarcha ce
990 6 1.5 224 5 Q11003 011003 schistosoma
991 6 1.5 224 8 Q9B116 09b116 nesotes con
992 6 1.5 224 8 Q9B115 09b115 nesotes fus
993 6 1.5 224 8 Q9B114 09b114 nesotes lin
994 6 1.5 224 8 Q9B113 09b113 nesotes pil
995 6 1.5 224 8 Q9B112 09b112 nesotes qua
996 6 1.5 224 8 Q9B120 09b120 nesotes qua
997 6 1.5 224 8 Q9B122 09b122 nesotes aet
998 6 1.5 224 8 Q9B121 09b121 nesotes con
999 6 1.5 224 8 Q94ZU6 094zu6 synemon sel
1000 6 1.5 224 8 Q94ZU5 094zu5 synemon par

```

## ALIGNMENTS

```

RESULT 1
Q9ZNM3 PRELIMINARY; PRT; 705 AA.
AC Q9ZNM3;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE iuta.
GN IUTA.
OS Vibrio orientalis.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=28175;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SD004;
RA Murakami K., Fuse H., Takimura O., Inoue H., Yamaoka Y.;
RT "Cloning and characterization of iuta gene from Vibrio species.";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB010890; BAA74703.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.
SQ SEQUENCE 705 AA; 77907 MW; 440CB4C074E44212 CRC64;

```

```

Query Match 3.2%; Score 13; DB 2; Length 705;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 138 RIEVLSGATSIYG 150
Db 132 RIEVLSGATSIYG 144

```

```

RESULT 2
Q934J5 PRELIMINARY; PRT; 725 AA.
AC Q934J5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE ferric aerobactin receptor (ferric aerobactin receptor precursor
GN iuta).
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=670;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WP1;
RA Funahashi T., Yamamoto S.;
RT "Cloning and characterization of the iuta gene in Vibrio
parahaemolyticus.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WP1;
RA Funahashi T., Yamamoto S.;
RT "Cloning and characterization of the iuta gene in Vibrio
parahaemolyticus.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB050700; BAB47148.1; -.
DR EMBL; AB066099; BAB83802.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR RECEPTOR.
SQ SEQUENCE 725 AA; 79329 MW; 57BDICE14CDA8748 CRC64;

```

```

Query Match 3.2%; Score 13; DB 2; Length 725;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 138 RIEVLSGATSIYG 150
Db 137 RIEVLSGATSIYG 149

```

```

RESULT 3
Q8ZHB0 PRELIMINARY; PRT; 726 AA.
AC Q8ZHB0;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative ferric siderophore receptor.
GN IUTA OR YPO0994.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Yersinia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / BIOVAR ORIENTALIS;
EX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Baeham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
DR EMBL; AJ414145; CAC89837.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.

```



KW Receptor; Hypothetical protein; Complete proteome.  
SQ SEQUENCE 726 AA; 80375 MW; 4EF2CA8010EF94C CRC64;

Query Match 2.5%; Score 10; DB 16; Length 726;  
Best Local Similarity 100.0%; Pred. No. 0.26;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 TGGGINIVTK 163  
Db 138 TGGGINIVTK 147

## RESULT 4

Q9XCH0 PRELIMINARY; PRT; 732 AA.  
ID Q9XCH0;  
AC Q9XCH0;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE IuA.  
GN IuA.  
OS Shigella flexneri.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Shigella.  
OX NCBI\_TaxID=623;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=M90T.  
RX MEDLINE=99340540; PubMed=10411725;  
RA Mose J.E., Cardozo T.J., Zychlinsky A., Groisman E.A.;  
RT "The *seC*-associated *SHI-2* pathogenicity island of *Shigella flexneri*.";  
RL Mol. Microbiol. 33:74-83(1999).  
DR EMBL: AF141323; AAD4750.1; -.  
DR InterPro: IPR000531; TonB\_boxC.  
DR Pfam: PF00593; TonB\_boxC.1.  
DR PROSITE: PS01156; TonB DEPENDENT REC\_2; UNKNOWN 1.  
SQ SEQUENCE 732 AA; 80984 MW; F3A1E0A1B7473210 CRC64;

Query Match 2.5%; Score 10; DB 2; Length 732;  
Best Local Similarity 100.0%; Pred. No. 0.26;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 TGGGINIVTK 163  
Db 143 TGGGINIVTK 152

## RESULT 5

Q93PE2 PRELIMINARY; PRT; 732 AA.  
ID Q93PE2;  
AC Q93PE2;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE IuA.  
GN IuA.  
OS Shigella boydii.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Shigella.  
OX NCBI\_TaxID=621;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=0-1392; PubMed=11418557;  
RX MEDLINE=21311742; PubMed=11418557;  
RA Purdy G.B., Payne S.M.;  
RT "The *SHI-3* Iron Transport Island of *Shigella boydii* 0-1392 Carries the  
RT Genes for Aerobactin Synthesis and Transport.";  
RL J. Bacteriol. 183:4176-4182(2001).  
DR EMBL: AF335540; AK71635.1; -.  
DR InterPro: IPR000531; TonB\_boxC.  
DR Pfam: PF00593; TonB\_boxC.1.  
DR PROSITE: PS01156; TonB DEPENDENT REC\_2; UNKNOWN 1.

SQ SEQUENCE 732 AA; 81013 MW; 57A3D10AEF70AE31 CRC64;

Query Match 2.5%; Score 10; DB 2; Length 732;  
Best Local Similarity 100.0%; Pred. No. 0.26;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 TGGGINIVTK 163  
Db 143 TGGGINIVTK 152

## RESULT 6

Q913X9 PRELIMINARY; PRT; 813 AA.  
ID Q913X9;  
AC Q913X9;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
DE Probable siderophore receptor.  
GN PA1365.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 15692 / PA01;  
RX MEDLINE=20437337; PubMed=10984043;  
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
RA Hickey M.J., Brinkman P.S.L., Huftagle W.O., Kowalik D.J., Lagrou M.,  
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
RA Brody L.L., Coulter S.N., Folger K.R., Kae A., Larbig K., Lim R.M.,  
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an  
RT opportunistic pathogen.";  
RL Nature 406:959-964(2000).  
DR EMBL: AE004565; AAG04754.1; -.  
DR InterPro: IPR000531; TonB\_boxC.  
DR Pfam: PF00593; TonB\_boxC.1.  
DR Receptor; Complete proteome.  
KW Receptor; Complete proteome.  
SQ SEQUENCE 813 AA; 89127 MW; 3E61AB56E3105299 CRC64;

Query Match 2.2%; Score 9; DB 16; Length 813;  
Best Local Similarity 100.0%; Pred. No. 3.4;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 IYSGGATGG 156  
Db 229 IYSGGATGG 237

## RESULT 7

O86513 PRELIMINARY; PRT; 334 AA.  
ID O86513;  
AC O86513;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Putative integral membrane protein.  
GN SC05531 OR SC1C2.12C.  
OS Streptomyces coelicolor.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RA Murphy L., Harris D.;  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);

```
RA Parkhill J., Barrell B.G., Rajandream M.A.;  
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RX MEDLINE=97000351; PubMed=8943436;  
RA Redenbach M., Kieser H.M., Denapite D., Eichner A., Cullum J.,  
RA Kinashi H., Hopwood D.A.;  
RT "A set of ordered cosmids and a detailed genetic and physical map for  
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";  
RN Mol. Microbiol. 21:77-96(1996).  
RN [4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,  
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
RT coelicolor A3(2).";  
RL Nature 417:141-147(2002).  
DR EMBL; AL031124; CAAL19979.1; -;  
DR InterPro; IPR000620; DUF6.  
DR Pfam; PF00892; DUF6; 2.  
SQ SEQUENCE 334 AA; 3495 MW; 718B756B50494EE CRC64;  
  
Query Match 2.0%; Score 8; DB 16; Length 334;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 90 LGVSSGTT 97  
Db 87 LGVSSGTT 94  
|||||  
  
RESULT 8  
Q8T293 PRELIMINARY; PRT; 335 AA.  
AC Q8T293;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Uncharacterized Fe-S oxidoreductase.  
GN MK0044.  
OS Methanopyrus kandleri.  
OC Archaea; Euryarchaeota; Methanopyri; Methanopyraceae;  
OC Methanopyrus.  
OX NCBI_TaxID=2320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AV19 / DSM 6324 / JCM 9639;  
RX MEDLINE=21927647; PubMed=11930014;  
RA Slesarev A.I., Mezhevaya K.V., Makarova K.S., Polushin N.N.,  
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,  
RA Natale D.A., Rogozin I.B., Tatusov R.L., Wolf Y.I., Stetter K.O.,  
RA Malykh A.G., Koonin E.V., Kozlyavkin S.A.;  
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19  
RT and monophyly of archaeal methanogens.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649(2002).  
DR EMBL; AE010305; AAM01261.1; -;  
DR Complete proteome.  
SQ SEQUENCE 335 AA; 38681 MW; F1519AFCC34FABEC CRC64;  
  
Query Match 2.0%; Score 8; DB 17; Length 335;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 327 AYRREKG 334
```

```
Db 13 AYRREKG 20  
|||||  
RESULT 9  
Q8XSQ1 PRELIMINARY; PRT; 812 AA.  
AC Q8XSQ1;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Probable ferric siderophore receptor outer membrane signal peptide  
DE protein.  
GN RSP0416 OR RS00873.  
OS Ralstonia solanacearum (Pseudomonas solanacearum).  
OC Plasmid megaplasmid.  
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;  
OC Ralstonia.  
OX NCBI_TaxID=305;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=GMI1000;  
RX MEDLINE=21681879; PubMed=11823852;  
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,  
RA Ariat M., Billault A., Brottier P., Camus J.-C., Cattolico L.,  
RA Chandler M., Choisme N., Claudel-Renard C., Cunnac S., Demange N.,  
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,  
RA Siguer P., Thebault P., Whalen M., Wincker P., Levy M.,  
RA Weissenbach J., Boucher C.A.;  
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";  
RL Nature 415:497-502(2002).  
DR EMBL; AL646078; CAD17567.1; -;  
DR InterPro; IPR000531; TonB_boxC.  
DR Pfam; PF00593; TonB_boxC; 1.  
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.  
KW Plasmid; Complete proteome.  
SQ SEQUENCE 812 AA; 87434 MW; DE03190282004BF3 CRC64;  
  
Query Match 2.0%; Score 8; DB 16; Length 812;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 149 YGSGATGG 156  
Db 228 YGSGATGG 235  
|||||  
  
RESULT 10  
Q9ULG1 PRELIMINARY; PRT; 1561 AA.  
AC Q9ULG1;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE KIAA1259 protein (Fragment).  
GN KIAA1259.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=BRAIN;  
RX MEDLINE=20039619; PubMed=10574462;  
RA Nagase T., Ishikawa K., Kikuno R., Hiroseawa M., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XV.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";  
RL DNA Res. 6:337-345(1999).  
DR EMBL; AB033085; BAA86573.1; -;  
DR InterPro; IPR001410; DEAD.  
DR InterPro; IPR001650; Helicase_C.  
DR InterPro; IPR000330; SNF2_N.
```

DR Pfam; PF00271; helicase\_C; 1.  
 DR Pfam; PF00176; SNF2\_N; 1.  
 DR SMART; SM00487; DEXDC; 1.  
 DR SMART; SM00490; HELICC; 1.  
 KM ATP-binding; Helicase.  
 PT NON\_TER  
 SQ SEQUENCE 1561 AA; 177340 MW; ABED6084A04E9C0B CRC64;

Query Match 2.0%; Score 8; DB 4; Length 1561;  
 Best Local Similarity 100.0%; Pred. No. 73;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 GYLQESK 382  
 DB 105 GYLQESK 112

## RESULT 11

ID Q8SSW5 PRELIMINARY; PRT; 2621 AA.  
 AC Q8SSW5;  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE KIAA1058 protein.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostelidae; Dictyostelium.  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX4;  
 RA Gloeckner G., Eichinger L., Szafranski K., Pachbat J., Dear P.,  
 RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,  
 RA Tunggal B., Cox B., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;  
 RT "Sequence and Analysis of Chromosome 2 of Dictyostelium."  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; A015585; AAL92252.1;  
 SQ SEQUENCE 2621 AA; 293740 MW; BF6A8B5C1B7DD210 CRC64;

Query Match 2.0%; Score 8; DB 5; Length 2621;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 PLTGSRD 124  
 DB 150 PLTGSRD 157

## RESULT 12

ID Q8Z5T7 PRELIMINARY; PRT; 63 AA.  
 AC Q8Z5T7;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Hypothetical protein STY2148.  
 GN STY2148.  
 OS Salmonella typhi.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Salmonella.  
 OX NCBI\_TaxID=601;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CT18;  
 RX MEDLINE=21534947; PubMed=11677608;  
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,  
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebathia M.,  
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,  
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
 RA Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagals K.,  
 RA Krogh A., Larsen T.S., Leach S., Moule S., O'Gaora P., Parry C.,  
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
 RA Whitehead S., Barrall B.G.;

RT "Complete genome sequence of a multiple drug resistant Salmonella  
 RT enterica serovar Typhi CT18."  
 RL Nature 413:848-852(2001).  
 DR EMBL; AL627272; CAD05690.1;  
 KM Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 63 AA; 7349 MW; AB319B8BCDE5284 CRC64;

Query Match 1.8%; Score 7; DB 16; Length 63;  
 Best Local Similarity 100.0%; Pred. No. 37;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 SLPLSV 17  
 DB 43 SLPLSV 49

## RESULT 13

ID Q9Y3H9 PRELIMINARY; PRT; 68 AA.  
 AC Q9Y3H9;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE DAL1H5.1b (Part of a chordin like protein with von Willebrand factor  
 DE type C domains) (Fragment).  
 GN DAL1H5.1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Pearce A.;  
 RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL049176; CAB42791.1;  
 DR InterPro; IPR001007; VWF\_C.  
 DR Pfam; PF00093; VWC; 1.  
 DR SMART; SM00214; VWC; 1.  
 DR PROSITE; PS01208; VWF\_C; UNKNOWN\_1.  
 FT NON\_TER  
 PT NON\_TER  
 SQ SEQUENCE 68 AA; 7918 MW; B97CC8ABA5F2285D CRC64;

Query Match 1.8%; Score 7; DB 4; Length 68;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 SENGAVL 205  
 DB 36 SENGAVL 42

## RESULT 14

ID Q9X9G0 PRELIMINARY; PRT; 79 AA.  
 AC Q9X9G0;  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE Hypothetical 9.2 kDa protein.  
 OS Yersinia pseudotuberculosis.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Yersinia.  
 OX NCBI\_TaxID=633;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9426823; PubMed=10496905;  
 RA Rakin A., Noelling C., Schubert S., Heesemann J.;  
 RT "Common and specific characteristics of the high-pathogenicity island  
 RL of Yersinia enterocolitica."  
 Infect. Immun. 67:5265-5274(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.

```

RA Rakin A., Schubert S., Gullvout I., Carniel E., Heesemann J.;
RT "Local hopping of 153 elements into the A+T-rich part of the high-
RT pathogenicity island in Yersinia enterocolitica 1B, O:8."
RL FEMS Microbiol. Lett. 182:2255-2294(2000).
DR EMBL; AJ236887; CAB46602.1; -.
KW Hypothetical protein.
SQ SEQUENCE 79 AA; 9199 MW; 62F484D3399D5F45 CRC64;

Query Match 1.8%; Score 7; DB 2; Length 79;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 298 LSEQPKT 304
Db 56 LSEQPKT 62

RESULT 15
Q9YGX0 PRELIMINARY; PRT; 83 AA.
AC Q9YGX0;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Short chain neurotoxin.
OS Laticauda colubrina (Yellow-lipped sea krait).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Elapidae; Laticaudinae; Laticauda.
OX NCBI_TaxID=8628;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Snake VENOM GLAND;
RA Kariya Y., Araki S., Agu H., Tamiya T., Tsuchiya T.;
RT "Classification of sea snakes in genus Laticauda by nucleotide
RT sequences encoding short chain neurotoxins."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017942; BAA75762.1; -.
DR HSP; P01435; IQKE.
DR InterPro; IPR003571; Snake_toxin.
DR Pfam; PF00087; toxin; 1.
DR ProDom; PD000206; Snake_toxin; 1.
DR PROSITE; PS00272; SNAKE_TOXIN; 1.
KW Neurotoxin.
SQ SEQUENCE 83 AA; 9343 MW; 1FE8212D22CA15D7 CRC64;

Query Match 1.8%; Score 7; DB 13; Length 83;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 301 QPKTTKS 307
Db 31 QPKTTKS 37

RESULT 16
Q9PWJ4 PRELIMINARY; PRT; 83 AA.
AC Q9PWJ4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Short chain neurotoxin.
OS Laticauda colubrina (Yellow-lipped sea krait).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Elapidae; Laticaudinae; Laticauda.
OX NCBI_TaxID=8628;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Snake VENOM GLAND;
RA Kariya Y., Araki S., Agu H., Tamiya T., Tsuchiya T.;
RT "Classification of sea snakes in genus Laticauda by nucleotide
RT sequences encoding short chain neurotoxins."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017942; BAA75762.1; -.
DR HSP; P01435; IQKE.
DR InterPro; IPR003571; Snake_toxin.
DR Pfam; PF00087; toxin; 1.
DR ProDom; PD000206; Snake_toxin; 1.
DR PROSITE; PS00272; SNAKE_TOXIN; 1.
KW Neurotoxin.
SQ SEQUENCE 83 AA; 9343 MW; 1FE8212D22CA15D7 CRC64;

Query Match 1.8%; Score 7; DB 13; Length 83;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 301 QPKTTKS 307
Db 31 QPKTTKS 37

RESULT 17
Q9YGC7 PRELIMINARY; PRT; 83 AA.
AC Q9YGC7;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Short chain neurotoxin.
OS Laticauda colubrina (Yellow-lipped sea krait).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Elapidae; Laticaudinae; Laticauda.
OX NCBI_TaxID=8628;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Snake VENOM GLAND;
RA Kariya Y., Araki S., Agu H., Tamiya T., Tsuchiya T.;
RT "Classification of sea snakes in genus Laticauda by nucleotide
RT sequences encoding short chain neurotoxins."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017944; BAA75764.1; -.
DR EMBL; AB017940; BAA75760.1; -.
DR EMBL; AB017943; BAA75763.1; -.
DR HSP; P01435; IQKE.
DR InterPro; IPR003571; Snake_toxin.
DR Pfam; PF00087; toxin; 1.
DR ProDom; PD000206; Snake_toxin; 1.
DR PROSITE; PS00272; SNAKE_TOXIN; 1.
KW Neurotoxin.
SQ SEQUENCE 83 AA; 9373 MW; 32F7031022CA02F0 CRC64;

Query Match 1.8%; Score 7; DB 13; Length 83;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 301 QPKTTKS 307
Db 31 QPKTTKS 37

RESULT 18
Q9PRJ7 PRELIMINARY; PRT; 83 AA.
AC Q9PRJ7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Short chain neurotoxin.
OS Laticauda colubrina (Yellow-lipped sea krait).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Elapidae; Laticaudinae; Laticauda.
OX NCBI_TaxID=8628;

```

```

RT sequences encoding short chain neurotoxins."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017947; BAA75767.1; -.
DR HSP; P01435; IQKE.
DR InterPro; IPR003571; Snake_toxin.
DR Pfam; PF00087; toxin; 1.
DR ProDom; PD000206; Snake_toxin; 1.
DR PROSITE; PS00272; SNAKE_TOXIN; 1.
KW Neurotoxin.
SQ SEQUENCE 83 AA; 9387 MW; 3752BC105C18B1F0 CRC64;

Query Match 1.8%; Score 7; DB 13; Length 83;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 301 QPKTTKS 307
Db 31 QPKTTKS 37

RESULT 17
Q9YGC7 PRELIMINARY; PRT; 83 AA.
AC Q9YGC7;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Short chain neurotoxin.
OS Laticauda colubrina (Yellow-lipped sea krait).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Elapidae; Laticaudinae; Laticauda.
OX NCBI_TaxID=8628;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Snake VENOM GLAND;
RA Kariya Y., Araki S., Agu H., Tamiya T., Tsuchiya T.;
RT "Classification of sea snakes in genus Laticauda by nucleotide
RT sequences encoding short chain neurotoxins."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017944; BAA75764.1; -.
DR EMBL; AB017940; BAA75760.1; -.
DR EMBL; AB017943; BAA75763.1; -.
DR HSP; P01435; IQKE.
DR InterPro; IPR003571; Snake_toxin.
DR Pfam; PF00087; toxin; 1.
DR ProDom; PD000206; Snake_toxin; 1.
DR PROSITE; PS00272; SNAKE_TOXIN; 1.
KW Neurotoxin.
SQ SEQUENCE 83 AA; 9373 MW; 32F7031022CA02F0 CRC64;

Query Match 1.8%; Score 7; DB 13; Length 83;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 301 QPKTTKS 307
Db 31 QPKTTKS 37

RESULT 18
Q9PRJ7 PRELIMINARY; PRT; 83 AA.
AC Q9PRJ7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Short chain neurotoxin.
OS Laticauda colubrina (Yellow-lipped sea krait).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Elapidae; Laticaudinae; Laticauda.
OX NCBI_TaxID=8628;

```

```

RN [1]
  SEQUENCE FROM N.A.
RC TISSUE=SNKE VENOM GLAND;
RA Kariya Y., Araki S., Agu H., Tamiya T., Tsuchiya T.;
RT "Classification of sea snakes in genus Laticauda by nucleotide
  sequences encoding short chain neurotoxins."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017941; BAA75761.1; -
DR EMBL; AB017937; BAA75757.1; -
DR EMBL; AB017938; BAA75758.1; -
DR EMBL; AB017939; BAA75759.1; -
DR HSSP; P01435; IOKE.
DR InterPro; IPR003571; Snake_toxin.
DR Pfam; PF00087; toxin; 1.
DR ProDom; PD000206; Snake_toxin; 1.
DR PROSITE; PS00272; SNAKE_TOXIN; 1.
KW Neurotoxin.
SQ SEQUENCE 83 AA; 9346 MW; 32E170C5F9B8B1F0 CRC64;

```

```

Query Match 1.8%; Score 7; DB 13; Length 83;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 301 QPKTKS 307
DB 31 QPKTKS 37

```

```

RESULT 19
QPRJ06 PRELIMINARY; PRT; 83 AA.
AC QPRJ06;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE Short chain neurotoxin.
OS Laticauda colubrina (yellow-lipped sea krait).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Laticaudinae; Laticauda.
OX NCBI_TaxID=8628;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SNKE VENOM GLAND;
RA Kariya Y., Araki S., Agu H., Tamiya T., Tsuchiya T.;
RT "Classification of sea snakes in genus Laticauda by nucleotide
  sequences encoding short chain neurotoxins."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017946; BAA75766.1; -
DR EMBL; AB017945; BAA75765.1; -
DR HSSP; P01435; IOKE.
DR InterPro; IPR003571; Snake_toxin.
DR Pfam; PF00087; toxin; 1.
DR ProDom; PD000206; Snake_toxin; 1.
DR PROSITE; PS00272; SNAKE_TOXIN; 1.
KW Neurotoxin.
SQ SEQUENCE 83 AA; 9357 MW; 32F7030AF9A7AEF0 CRC64;

```

```

Query Match 1.8%; Score 7; DB 13; Length 83;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 301 QPKTKS 307
DB 31 QPKTKS 37

```

```

RESULT 20
QPRJ05 PRELIMINARY; PRT; 83 AA.
AC QPRJ05;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)

```

```

DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE Short chain neurotoxin.
OS Laticauda colubrina (yellow-lipped sea krait).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Laticaudinae; Laticauda.
OX NCBI_TaxID=8628;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SNKE VENOM GLAND;
RA Kariya Y., Araki S., Agu H., Tamiya T., Tsuchiya T.;
RT "Classification of sea snakes in genus Laticauda by nucleotide
  sequences encoding short chain neurotoxins."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017954; BAA75774.1; -
DR EMBL; AB017948; BAA75768.1; -
DR EMBL; AB017949; BAA75769.1; -
DR EMBL; AB017950; BAA75770.1; -
DR EMBL; AB017951; BAA75771.1; -
DR EMBL; AB017952; BAA75772.1; -
DR EMBL; AB017953; BAA75773.1; -
DR HSSP; P01435; IOKE.
DR InterPro; IPR003571; Snake_toxin.
DR Pfam; PF00087; toxin; 1.
DR ProDom; PD000206; Snake_toxin; 1.
DR PROSITE; PS00272; SNAKE_TOXIN; 1.
KW Neurotoxin.
SQ SEQUENCE 83 AA; 9373 MW; 32F71C15F9B8B1F0 CRC64;

```

```

Query Match 1.8%; Score 7; DB 13; Length 83;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 301 QPKTKS 307
DB 31 QPKTKS 37

```

```

RESULT 21
QPRJ00 PRELIMINARY; PRT; 83 AA.
AC QPRJ00;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE Short chain neurotoxin.
OS Laticauda colubrina (yellow-lipped sea krait).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Elapidae; Laticaudinae; Laticauda.
OX NCBI_TaxID=8628;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SNKE VENOM GLAND;
RA Kariya Y., Araki S., Agu H., Tamiya T., Tsuchiya T.;
RT "Classification of sea snakes in genus Laticauda by nucleotide
  sequences encoding short chain neurotoxins."
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB017936; BAA75756.1; -
DR EMBL; AB017933; BAA75753.1; -
DR EMBL; AB017934; BAA75754.1; -
DR EMBL; AB017935; BAA75755.1; -
DR HSSP; P01435; IOKE.
DR InterPro; IPR003571; Snake_toxin.
DR Pfam; PF00087; toxin; 1.
DR ProDom; PD000206; Snake_toxin; 1.
DR PROSITE; PS00272; SNAKE_TOXIN; 1.
KW Neurotoxin.
SQ SEQUENCE 83 AA; 9314 MW; 45FCDA72F9B8A5B4 CRC64;

```

```

Query Match 1.8%; Score 7; DB 13; Length 83;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 301 QPKTKS 307  
|  
|  
|  
|  
|  
Db 31 QPKTKS 37

## RESULT 22

ID Q8VKB3 PRELIMINARY; PRT; 85 AA.  
AC Q8VKB3;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Hypothetical protein MT0943.  
GN MT0943.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / OGHKOSH;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E., S.L.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W.;  
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
laboratory strains";  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE006980; AAK45190.1; -;  
DR TIGR; MT0943; -;  
KW Hypothetical protein.  
SQ SEQUENCE 85 AA; 8533 MW; ADF2CA6D275E0C1A CRC64;

Query Match 1.8%; Score 7; DB 16; Length 85;  
Best Local Similarity 100.0%; Pred. No. 49;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 115 GVPLTGS 121  
|  
|  
|  
|  
|  
Db 57 GVPLTGS 63

## RESULT 23

ID Q9CJE2 PRELIMINARY; PRT; 87 AA.  
AC Q9CJE2;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Hypothetical protein yafJ.  
GN YAFJ OR LJ0055.  
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;  
OC Streptococcaceae; Lactococcus.  
OX NCBI\_TaxID=13360;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IL1403;  
RX MEDLINE=21235186; PubMed=11337471;  
RA Bolotin A., Wincker P., Manger S., Jaillon O., Malarne K.,  
RA Weissenbach J., Ehrlich S.D., Sorokin A.;  
RT "The complete genome sequence of the lactic acid bacterium Lactococcus  
lactis ssp. lactis IL1403";  
RT Genome Res. 11:731-753(2001).  
RL EMBL; AE006244; AAK04153.1; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 87 AA; 9014 MW; DD24D5C6A2322836 CRC64;

Query Match 1.8%; Score 7; DB 16; Length 87;  
Best Local Similarity 100.0%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 91 GVSSGTT 97  
|  
|  
|  
|  
|  
Db 43 GVSSGTT 49

## RESULT 24

ID Q9LWDO PRELIMINARY; PRT; 95 AA.  
AC Q9LWDO;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE ESTs C27640(C52458).  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. NIPPONBARE;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
clone: P0483F08.";  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP002094; BAA96214.1; -;  
SQ SEQUENCE 95 AA; 10855 MW; AC2B618BB9F71A7F CRC64;

Query Match 1.8%; Score 7; DB 10; Length 95;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 359 SATLDAY 365  
|  
|  
|  
|  
|  
Db 31 SATLDAY 37

## RESULT 25

ID Q9M581 PRELIMINARY; PRT; 101 AA.  
AC Q9M581;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Plastid-specific ribosomal protein 4 precursor.  
GN PSRP-4.  
OS Spinacia oleracea (Spinach).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Spinacia.  
OX NCBI\_TaxID=3562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. ALVARO;  
RA Yanaguchi K., von Knoblauch K., Subramanian A.R.;  
RT "PSRP-4: an unusual chloroplast ribosomal protein with homology to a  
Thermus thermophilus protein (and no other bacterial proteins in  
databases) and a putative plant mitochondrial protein.";  
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF236825; AAF64154.1; -;  
KW Ribosomal protein; Transit peptide.  
FT TRANSIT 1 54  
FT CHAIN 55 101 PLASTID-SPECIFIC RIBOSOMAL PROTEIN 4.  
SQ SEQUENCE 101 AA; 10585 MW; E93AGB76509414C3 CRC64;

Query Match 1.8%; Score 7; DB 10; Length 101;  
Best Local Similarity 100.0%; Pred. No. 58;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 193 QSVAGVS 199  
|  
|  
|  
|  
|  
Db 24 QSVAGVS 30

```

RESULT 26
O9DM6 PRELIMINARY; PRT; 108 AA.
AC O9DM6;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE 1300002P07Rik protein.
GN 1300002P07Rik
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
RA Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matuda H. A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Glisic C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L. M., Staahl F., Suzuki R., Tomita M., Wagner L., Maehio T.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M. F.,
RA Brownstein M. J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustinck S., Hill D., Hofmann M., Hume D. A., Kamiya M., Lee N. H.,
RA Lyons P., Marchionni L., Mashima M., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Saito K., Schoenbach C., Seya T., Shibata Y., Storch K. -F.,
RA Suzuki H., Tovo-Oka K., Wang K. H., Weitz C., Whitaker C., Wilmink L.,
RA Wyszewski B. A., Yoshida K., Hasegawa Y., Kawai H., Kohseki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
CC -1- COFACTOR: ZINC (BY SIMILARITY).
DR EMBL; AK004863; BAB23624.1; -
DR MGD; MGI:1918999; 1300002P07Rik.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; ADH_zn_family.
DR Pfam; PF00107; adh_zinc; 1.
DR PROSITE; PS00059; ADH_ZINC; 1.
KW Oxidoreductase; Zinc.
SQ SEQUENCE 108 AA; 11696 MW; 345D0763E89621F7 CRC64;

Query Match 1.8%; Score 7; DB 11; Length 108;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 70 QEOATGS 76
DB 93 QEOATGS 99

RESULT 27
O8ZNU2 PRELIMINARY; PRT; 108 AA.
ID O8ZNU2;
AC O8ZNU2;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DE Putative glucose-6-phosphate dehydrogenase.
GN STM1939.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxId=602;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K. E., Spieth J., Clifton S. W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stonking T., Nhan M.,
RA Waterston R., Wilson R. K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856 (2001).
DR EMBL; AE008786; ALU20854.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 108 AA; 12649 MW; 4DDBDE3B7185DEF8 CRC64;

Query Match 1.8%; Score 7; DB 16; Length 108;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 11 SUPLSV 17
DB 43 SUPLSV 49

RESULT 28
O9PXU5 PRELIMINARY; PRT; 115 AA.
ID O9PXU5;
AC O9PXU5;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Coat protein (Fragment).
OS Sugarcane mosaic virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Polyviridae;
OC Potyvir.
OX NCBI_TaxId=12224;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93393440; PubMed=8379856;
RA Xiao X. W., Frenkel M. J., Teakle D. S., Ward C. W., Shukla D. D.;
RT "Sequence diversity in the surface-exposed amino-terminal region of
RT the coat proteins of seven strains of sugarcane mosaic virus
RT correlates with their host range."
RL Arch. Virol. 132:399-408 (1993).
DR InterPro; IPR001670; Fe-ADH.
DR InterPro; IPR001592; Poty_coat.
DR Pfam; PF00767; Poty_coat; 1.
DR PRINTS; PRO1574; TUBBYPROTEIN.
DR PROSITE; PS00060; ADH_IRON_2; UNKNOWN 1.
SQ SEQUENCE 115 AA; 9736 MW; 89E18950DB527A21 CRC64;

Query Match 1.8%; Score 7; DB 12; Length 115;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 150 GSGATCG 156
DB 66 GSGATCG 72

RESULT 29
O9WH59 PRELIMINARY; PRT; 130 AA.
ID O9WH59;
AC O9WH59;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Mitochondrial protein (Fragment).
GN RPS2.
OS Ginkgo biloba (Ginkgo).
OC Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Ginkgophyta; Ginkgoales; Ginkgoaceae; Ginkgo.
OX NCBI_TaxId=3311;

```

```

RN RP SEQUENCE FROM N.A.
RA Parkinson C.L., Adams K.L., Palmer J.D.;
RT "Multigene analyses identify the three earliest lineages of extant
RL flowering plants.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF193904; AAF77639.1; -.
DR InterPro; IP0001865; Ribosomal_S2.
DR PROSITE; PS00963; RIBOSOMAL_S2_2; UNKNOWN_1.
KW Mitochondrion.
FT NON_TER 1
FT NON_TER 130 130
SQ SEQUENCE 130 AA; 1492 MW; 26136D9ED921F995 CRC64;

Query Match 1.8%; Score 7; DB 8; Length 130;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 240 SLSVNTN 246
Db 15 SLSVNTN 21

RESULT 30
Q45435 PRELIMINARY; PRT; 154 AA.
AC Q45435;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE MnxB.
GN MnxB.
OS Bacillus sp.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=1409;
RN RP SEQUENCE FROM N.A.
RA van Waasbergen L.G., Hildebrand M., Tebo B.M.;
RX MEDLINE=96256605; PubMed=8655549;
RT "Identification and characterization of a gene cluster involved in
RT manganese oxidation by spores of the marine Bacillus sp. strain SG-
RT 1.";
RL J. Bacteriol. 178:3517-3530(1996).
RN RP SEQUENCE FROM N.A.
RA van Waasbergen L.G., Tebo B.M.;
RX STRAIN=SG-1;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U31081; AAB06484.1; -.
SQ SEQUENCE 154 AA; 16465 MW; 8209D9ABEBB44540 CRC64;

Query Match 1.8%; Score 7; DB 2; Length 154;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 SGATGGL 157
Db 95 SGATGGL 101

RESULT 31
Q8QRX5 PRELIMINARY; PRT; 156 AA.
AC Q8QRX5;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE UL145.
OS chimpanzee cytomegalovirus.
OC viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.

```

```

OX NCBI_TaxID=188763;
RN RP SEQUENCE FROM N.A.
RA Davison A.J., Akter P., Dolan A., Wright K.M., Addison C.,
RA Alcendor D.J., Hayward G.S., McGeoch D.J.;
RT "The human cytomegalovirus genome revisited.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF480884; AAM00762.1; -.
SQ SEQUENCE 156 AA; 17040 MW; 8AC585437190BC3C CRC64;

Query Match 1.8%; Score 7; DB 12; Length 156;
Best Local Similarity 100.0%; Pred. No. 89;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 203 NVLARLD 209
Db 92 NVLARLD 98

RESULT 32
Q90494 PRELIMINARY; PRT; 161 AA.
ID Q90494;
AC Q90494;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Vasotocin precursor.
GN PREPRO-VASOTOCIN.
OS Eptatretus burgeri (Inshore hagfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Hyperotreti; Myxiniiformes;
OC Myxiniidae; Eptatretinae; Eptatretus.
OX NCBI_TaxID=7764;
RN RP SEQUENCE FROM N.A.
RA Suzuki M., Kubokawa K., Nagasawa H., Urano A.;
RC TISSUE=BRAIN;
RT "Sequence analysis of vasotocin cDNAs of the lamprey, Lampetra
RT japonica, and the hagfish, Eptatretus burgeri: Evolution of cyclostome
RT vasotocin precursors.";
RL J. Mol. Endocrinol. 0:0-0(1994).
DR EMBL; D31870; BAA06668.1; -.
DR HSPF; P01180; INPO.
DR InterPro; IP000981; Neurhyp_horm.
DR Pfam; PF00220; hormones_1.
DR Pfam; PF00184; hormones_1.
DR ProDom; PD001676; Neurhyp_horm; 1.
DR SMART; SM00003; NH; 1.
DR PROSITE; PS00264; NEUROHYPOPHYS_HORM; 1.
KW Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 31 VASOTOCIN.
FT CHAIN 35 161 NEUROPHYSIN.
SQ SEQUENCE 161 AA; 15886 MW; 2CD388F2E3ABDAE5 CRC64;

Query Match 1.8%; Score 7; DB 13; Length 161;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 151 SGATGGL 157
Db 135 SGATGGL 141

RESULT 33
Q9HNS0 PRELIMINARY; PRT; 173 AA.
ID Q9HNS0;
AC Q9HNS0;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE Vng1974h.
GN Vng1974h.
OS Halobacterium sp. (strain NRC-1).

```



```

OC Archaeae; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Halobacterium.
OK NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berguist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsen V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Weir R., Goo Y.A.,
RA Leithauer B., Keller K., Cruz R., Hanson M.J., Hough D.W.,
RA Maddock D.G., Jablonki P.E., Krebs M.P., Angermeier C.M., Dale H.,
RA Isebaeager T.A., Peck R.F., Pohlshroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;
RA "Genome sequence of Halobacterium species NRC-1."
RT Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
RL EMBL; AE005093; MAG20150.1; -.
DR Complete proteome.
SQ SEQUENCE 173 AA; 17801 MW; F65E6P9A3323F341 CRC64;

Query Match 1.8%; Score 7; DB 17; Length 173;
Best Local Similarity 100.0%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 GGAFFDAN 222
DB 118 GGAFFDAN 124

RESULT 34
ID Q9F3Q2 PRELIMINARY; PRT; 182 AA.
AC Q9F3Q2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Secreted extracellular small neutral protease.
GN SNP3 OR SCO7635 OR SC10F4.08C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OK NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinovitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorek A., Woodward J., Barrell B.G., Parkhill J.,
Hopwood D.A.;

```

```

RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2)."
RT Nature 417:141-147(2002).
RL EMBL; AL450350; CAC16968.1; -.
DR HSSP; P56406; 1KTH.
DR InterPro; IPR000013; Peptidase_M7.
DR InterPro; IPR000130; Zn_Mpeptidase.
DR Pfam; PF02031; Peptidase_M7_1.
DR PRINTS; PR00787; NEUTRALPTASE.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Protease.
SQ SEQUENCE 182 AA; 18770 MW; F1E1E339A4F83CC1 CRC64;

Query Match 1.8%; Score 7; DB 16; Length 182;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 LSLPLIS 16
DB 15 LSLPLIS 21

RESULT 35
ID Q9VR84 PRELIMINARY; PRT; 183 AA.
AC Q9VR84;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE CG1631 protein.
GN CG1631.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OK NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinkner S.B., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertan B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jaleel M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pauley J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Relezz K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kimios I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector A., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.W., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

```

```

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003571; AAF50920.1; -.
DR FlyBase; FBgn0031101; CG1631.
SQ SEQUENCE 183 AA; 20435 MW; EGE141F795E577543 CRC64;

Query Match 1.8%; Score 7; DB 5; Length 183;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 164 SDLEEEQ 170
Db 65 SDLEEEQ 71

RESULT 36
Q9QZJ2 PRELIMINARY; PRT; 187 AA.
AC Q9QZJ2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Na-taurocholate cotransporting polypeptide (Fragment).
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RA Balasubramanian N., Arrese M., Suchy F.J., Ananthanarayanan M.;
RT "Na-Taurocholate cotransporting polypeptide (Ntcp) from Hamster
liver.";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF181258; AAD3961.1; -.
DR InterPro; IPR002657; BileAc/Na_smptr.
DR Pfam; PF01758; SBP; 1.
FT NON_TER 1 187
FT NON_TER 187 187
SQ SEQUENCE 187 AA; 20182 MW; 2855C5F44AB482C6 CRC64;

Query Match 1.8%; Score 7; DB 11; Length 187;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 LLSVAVT 20
Db 130 LLSVAVT 136

RESULT 37
O61000 PRELIMINARY; PRT; 188 AA.
ID O61000;
AC O61000;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Tryparedoxin peroxidase.
GN TXNIPX.
OS Crithidia fasciculata.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Crithidia.
OX NCBI_TaxID=5656;
RN [1]
RP SEQUENCE FROM N.A.
RA Nogoceke E., Gommel D.U., Kiess M., Kalisz H.M., Flohe L.;
RX MEDLINE=98020808; PubMed=9377478;
RT "A unique cascade of oxidoreductases catalyzes trypanothione-mediated
peroxide metabolism in Crithidia fasciculata.";
RL Biol. Chem. 378:827-836(1997).

```

```

[2]
RP SEQUENCE FROM N.A.
RC STRAIN=HS6;
RX MEDLINE=98148021; PubMed=9478927;
RA Montemartini M., Nogoceke E., Singh M., Steinert P., Flohe L.,
RA Kalisz H.M.;
RT "Sequence analysis of the tryparedoxin peroxidase gene from Crithidia
fasciculata and its functional expression in Escherichia coli.";
RL J. Biol. Chem. 273:4864-4871(1998).
DR EMBL; AF020947; AAC15095.1; -.
DR HSSP; P30041; 1PRX.
DR InterPro; IPR000866; AhpC-TSA.
DR Pfam; PF00578; AhpC-TSA; 1.
KW Peroxidase.
SQ SEQUENCE 188 AA; 20944 MW; C6867035935A3142 CRC64;

Query Match 1.8%; Score 7; DB 5; Length 188;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 371 ARAYGVL 377
Db 113 ARAYGVL 119

RESULT 38
Q9TZX2 PRELIMINARY; PRT; 188 AA.
ID Q9TZX2;
AC Q9TZX2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Tryparedoxin peroxidase.
GN TRYP.
OS Crithidia fasciculata.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Crithidia.
OX NCBI_TaxID=5656;
RN [1]
RP SEQUENCE FROM N.A.
RA Tetaud E., Fairlamb A.H.;
RX MEDLINE=99066936; PubMed=9851611;
RT "Cloning, expression and reconstitution of the trypanothione-dependent
peroxidase system of Crithidia fasciculata.";
RL Mol. Biochem. Parasitol. 96:111-123(1998).
DR EMBL; AF055914; AAC72300.1; -.
DR HSSP; P30041; 1PRX.
DR InterPro; IPR000866; AhpC-TSA.
DR Pfam; PF00578; AhpC-TSA; 1.
KW Peroxidase.
SQ SEQUENCE 188 AA; 20958 MW; C2598B15935DA8DC CRC64;

Query Match 1.8%; Score 7; DB 5; Length 188;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 371 ARAYGVL 377
Db 113 ARAYGVL 119

RESULT 39
Q9GQW7 PRELIMINARY; PRT; 190 AA.
ID Q9GQW7;
AC Q9GQW7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Peroxidoxin 1.
OS Leishmania chagasi.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=44271;
RN [1]
RP SEQUENCE FROM N.A.

```

RX MEDLINE=21423979; PubMed=11438539;  
 RA Barr S.D., Gedamu L.;  
 RT "Cloning and Characterization of Three Differentially Expressed  
 RT Peroxidoxin Genes from Leishmania chagasi. EVIDENCE FOR AN ENZYMATIC  
 RT DETOXIFICATION OF HYDROXYL RADICALS."  
 RL J. Biol. Chem. 276:34279-34287(2001).  
 DR EMBL, AF134161; AAK40074.1; -;  
 DR InterPro; IPR000866; Ahpc-TSA.  
 DR Pfam; PF00578; Ahpc-TSA; 1.  
 SQ SEQUENCE 190 AA; 21280 MW; 9388978CBCE165B35 CRC64;

Query Match 1.8%; Score 7; DB 5; Length 190;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 ARAYGVL 377  
 DB 113 ARAYGVL 119

RESULT 40  
 Q95WZ8 PRELIMINARY; PRT; 190 AA.

AC Q95WZ8; 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Peroxidoxin 1.  
 GN PXN1.

OS Leishmania donovani.  
 OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OK NCBI\_TaxID=5661;  
 RP SEQUENCE FROM N.A.

RA Barr S.D., Gedamu L.;  
 RC MEDLINE=21423979; PubMed=11438539;  
 RT "Cloning and Characterization of Three Differentially Expressed  
 RT Peroxidoxin Genes from Leishmania chagasi. EVIDENCE FOR AN ENZYMATIC  
 RT DETOXIFICATION OF HYDROXYL RADICALS."  
 RL J. Biol. Chem. 276:34279-34287(2001).  
 DR EMBL, AF205887; AAK82654.1; -;  
 DR InterPro; IPR000866; Ahpc-TSA.  
 DR Pfam; PF00578; Ahpc-TSA; 1.  
 SQ SEQUENCE 190 AA; 21296 MW; 4D376798F9109F7D CRC64;

Query Match 1.8%; Score 7; DB 5; Length 190;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 ARAYGVL 377  
 DB 113 ARAYGVL 119

RESULT 41  
 Q9BP39 PRELIMINARY; PRT; 199 AA.

AC Q9BP39; 01-UN-2001 (TREMBLrel. 17, Created)  
 DT 01-UN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-UN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Tryparedoxin peroxidase.  
 GN TXNPX.

OS Leishmania donovani.  
 OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OK NCBI\_TaxID=5661;  
 RP SEQUENCE FROM N.A.

RA Karsel, Kalavari S., Steinert P., Singh M., Plohe L.;  
 RT "Leishmania donovani tryparedoxin peroxidase."  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL, AF225212; AAK0633.1; -;  
 DR InterPro; IPR000866; Ahpc-TSA.  
 DR Pfam; PF00578; Ahpc-TSA; 1.  
 KW Peroxidase.  
 SQ SEQUENCE 199 AA; 22195 MW; 6A6904E87979557 CRC64;

Query Match 1.8%; Score 7; DB 5; Length 199;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 ARAYGVL 377  
 DB 113 ARAYGVL 119

RESULT 42  
 Q95WL4 PRELIMINARY; PRT; 199 AA.

AC Q95WL4; 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Peroxidoxin 3.  
 GN PXN3.

OS Leishmania chagasi.  
 OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OK NCBI\_TaxID=44271;  
 RP SEQUENCE FROM N.A.

RA Barr S.D., Gedamu L.;  
 RC MEDLINE=21423979; PubMed=11438539;  
 RT "Cloning and Characterization of Three Differentially Expressed  
 RT Peroxidoxin Genes from Leishmania chagasi. EVIDENCE FOR AN ENZYMATIC  
 RT DETOXIFICATION OF HYDROXYL RADICALS."  
 RL J. Biol. Chem. 276:34279-34287(2001).  
 DR EMBL, AF312398; AAK69587.1; -;  
 DR InterPro; IPR000866; Ahpc-TSA.  
 DR Pfam; PF00578; Ahpc-TSA; 1.  
 SQ SEQUENCE 199 AA; 22123 MW; E33CD68CCD1D311E CRC64;

Query Match 1.8%; Score 7; DB 5; Length 199;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 ARAYGVL 377  
 DB 113 ARAYGVL 119

RESULT 43  
 Q95NFS PRELIMINARY; PRT; 199 AA.

AC Q95NFS; 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Cytosolic peroxidoxin (Peroxidoxin 2).  
 GN PXN2.

OS Leishmania infantum, and  
 OS Leishmania chagasi.  
 OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.  
 OK NCBI\_TaxID=5671, 44271;  
 RP SEQUENCE FROM N.A.

RA Caetano H., Tomas A.M.;  
 RT "Leishmania infantum CTNPPX gene encoding a cytosolic peroxidoxin."  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

RA Barr S.D., Gedamu L.;  
 RC MEDLINE=21423979; PubMed=11438539;  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

RT "Cloning and Characterization of Three Differentially Expressed  
RT Peroxidoxin Genes from Leishmania chagasi. EVIDENCE FOR AN ENZYMATIC  
RT DETOXIFICATION OF HYDROXYL RADICALS.";  
RL J. Biol. Chem. 276:34279-34287(2001).

DR EMBL; AY058210; AAL25847.1; -

DR EMBL; AF312397; AAK69586.1; -

DR InterPro; IPR000866; AbpC-TSA.

DR Pfam; PF00578; AbpC-TSA; 1.

SQ SEQUENCE 199 AA; 22138 MW; F2CC6D68CC1D311E CRC64;

Query Match 1.8%; Score 7; DB 5; Length 199;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 371 ARAYGV 377

|||||

Db 113 ARAYGV 119

RESULT 44

Q9HVB2

ID Q9HVB2 PRELIMINARY; PRT; 200 AA.

AC Q9HVB2;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)

DE Hypothetical protein PA4683.

PA4683.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCBI\_TaxID=287;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 15692 / PA01;

RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,

RA Garber R.L., Golltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,

RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,

RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.,

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an

RT opportunistic pathogen.";

RL Nature 406:959-964(2000).

DR EMBL; AE004882; AAG08070.1; -

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 200 AA; 22685 MW; DB1E3CA1A32CE1BB CRC64;

Query Match 1.8%; Score 7; DB 16; Length 200;

Best Local Similarity 100.0%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 289 SLNAIKG 295

|||||

Db 181 SLNAIKG 187

RESULT 45

Q8V5K7

ID Q8V5K7 PRELIMINARY; PRT; 202 AA.

AC Q8V5K7;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE 23K endoprotease.

OS Possum adenovirus 1.

OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Atadenovirus.

OX NCBI\_TaxID=150098;

RN [1]

RP SEQUENCE FROM N.A.

RA Thomson D.M., Harrach B., Meers J.;

RT "Molecular phylogeny of possum adenovirus: a new atadenovirus.";

RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF338823; AAL73248.1; -

DR InterPro; IPR000855; Peptidase\_C5.

DR Pfam; PF00770; Peptidase\_C5; 1.

DR PRINTS; PR00703; ADVENDOPTASE.

DR ProDom; PD003705; Peptidase\_C5; 1.

KW Protease.

SQ SEQUENCE 202 AA; 23104 MW; F3A172F9CBBEB7EA CRC64;

Query Match

Best Local Similarity 100.0%; Score 7; DB 12; Length 202;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 89 SLGVSSG 95

|||||

Db 15 SLGVSSG 21

Search completed: December 25, 2002, 20:32:06

Job time : 55 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002, Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 25, 2002, 20:33:19 / Search time 2671 Seconds

(without alignments)  
4358.339 Million cell updates/sec

Title: US-09-889-746-2

Perfect score: 2047

Sequence: 1 MSHSHYFQWLSPLLSAVT.....STAEVLGRVPINKKRALE 400

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DBV=xlh  
-O=/cg2.1/USPTO.sp08989746/rnat 20122002.143749.14202/app.query.fasta.1.583  
-DB=GenEmbl -OPMT=fastcp -SUFFIX=rge -MINMATCH=0.1 -LOOFCU=0 -LOOEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human40.cdi -LIST=45  
-DOCFALIG=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=45 -MODE=LOCAL  
-OUTFMT=prc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US0989746 @CGN 1.1 1616 @rnat 20122002.143749.14202 -NCEPU=6 -ICPU=3  
-NO\_XLPXY -NO\_NMAP -LARGEOUTERY -NEG SCORES=0 -WAIT -LONGLOG -DEV TIMOUT=120  
-WARN TIMOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_scs.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vl.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_scs.\*  
28: em\_un.\*

29: em\_vl.\*  
30: em\_hgt\_hum.\*  
31: em\_hgt\_inv.\*  
32: em\_hgt\_other.\*  
33: em\_hgt\_mus.\*  
34: em\_hgt\_pln.\*  
35: em\_hgt\_rtd.\*  
36: em\_hgt\_mam.\*  
37: em\_hgt\_vrt.\*  
38: em\_sy.\*  
39: em\_hgt\_hum.\*  
40: em\_hgt\_mus.\*  
41: em\_hgt\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2047	100.0	1203	1 AX027672	AX027672 Sequence
2	2047	100.0	1203	6 AX027664	AX027664 Sequence
3	652	31.9	162976	2 AC098122	AC098122 Rattus no
4	527	25.7	11579	1 AE004565	AE004565 Pseudomon
5	503.5	24.6	12421	1 AE004881	AE004881 Pseudomon
6	497	24.3	203050	1 AL646078	AL646078 Ralstonia
7	495	24.2	2803	1 AB050700	AB050700 Vibrio pa
8	495	24.2	5567	1 AB066099	AB066099 Vibrio pa
9	491	24.0	13252	1 AE013940	AE013940 Yersinia
10	491	24.0	208050	1 AJ414145	AJ414145 Yersinia
11	481.5	23.5	36688	1 AY046057	AY046057 Escherich
12	472	23.1	20962	1 AF335540	AF335540 Shigella
13	472	23.1	23771	1 AF141323	AF141323 Shigella
14	463	22.6	3223	6 A07458	A07458 Nucleotide
15	457	22.3	3219	6 A07457	A07457 Nucleotide
16	453	22.1	2769	1 ECDP13RE	ECDP13RE
17	453	22.1	11334	1 AE007312	AE007312 Sinorhizo
18	453	22.1	11853	1 AP110737	AP110737 Sinorhizo
19	453	22.1	342650	1 AP003582	AP003582 Nostoc sp
20	448.5	21.9	133859	1 D90899	D90899 Synchocyst
21	441	21.5	2734	1 AB010890	AB010890 Vibrio or
22	432	21.1	3495	1 AB074151	AB074151 Vibrio vu
23	400.5	19.6	341880	1 AP003589	AP003589 Nostoc sp
24	381.5	18.6	4290	1 AENLEBCTM	X97499 Ralstonia s
25	333.5	16.3	333500	1 AP003590	AP003590 Nostoc sp
26	326.5	16.0	334520	1 AP003588	AP003588 Nostoc sp
27	302	14.8	110000	2 LMFLCHR36_27	Continuation (28 o
28	270	13.2	110000	2 LMFLCHR32_10	Continuation (11 o
29	209.5	10.2	12514	1 AE012011	AE012011 Xanthomon
30	209	10.2	2050	1 PGH87395	PGH87395 Porphyromon
31	204.5	10.0	10233	1 AE010499	AE010499 Fusobacte
32	202.5	9.9	24454	1 AF048749	AF048749 Bacteroid
33	200	9.8	274050	1 AL627276	AL627276 Salmonell
34	199.5	9.7	2495	1 AF135597	AF135597 Escherich
35	199.5	9.7	4837	1 SEU97227	SEU97227 Salmonella
36	198	9.7	204050	1 AL646070	AL646070 Ralstonia
37	196.5	9.6	37084	1 ECF5FAGSH	X16664 Escherichia
38	195	9.5	9848	1 AY029471	AY029471 Salmonell
39	194	9.5	20513	1 AE008826	AE008826 Salmonell
40	193	9.4	10118	1 AE012418	AE012418 Xanthomon
41	190.5	9.3	10835	1 AE012012	AE012012 Xanthomon
42	189.5	9.3	10399	1 AY008342	AY008342 Plesiomon
43	189	9.2	10153	1 AE011991	AE011991 Xanthomon
44	187.5	9.2	10417	1 AE011961	AE011961 Xanthomon
45	187.5	9.2	348250	1 AP003592	AP003592 Nostoc sp

RESULT 1

## ALIGNMENTS

```

AX027672      AX027672      1203 bp      DNA      linear      BCT 16-SEP-2000
LOCUS
DEFINITION    Sequence 1 from Patent WO0043519.
ACCESSION    AX027672
VERSION      AX027672.1      GI:10188548
KEYWORDS
SOURCE      Neisseria meningitidis.
ORGANISM    Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE    1 (bases 1 to 1203)
AUTHORS      Ruelle,J.L.
TITLE        Novel compounds
JOURNAL      Patent: WO 0043519-A 27-JUL-2000;
            RUELE JEAN LOUIS (BE) ; SMITHKLINE BEECHAM BIOLOG (BE)
FEATURES
source
BASE COUNT    359 a 261 c 263 g 320 t
ORIGIN
Alignment Scores:
Pred. No.:    2,94e-142      Length:    1203
Score:        2047.00      Matches:    400
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 100.00%      Indels: 0
Gaps: 0
DB: 1
US-09-889-746-2 (1-400) x AX027672 (1-1203)
Qy 1 MetArgHisSerHisTyrPheGlnTrpLeuSerLeuProLeuLeuSerValAlaValThr 20
Db 1 ATGCGTCATTCCCACTATTTCAATGGTTATCTTTGCCCTTTACTAAGTGTGGCAGTAAT 60
Qy 21 GlnGlnLeuTyrAlaGlnProGlnGluSerLeuProThrValGluLeuGluProValVal 40
Db 61 CAGCAGTTGTAGCTCAACCCCAATGAGTCATTACCACCGGTTGAATTAGAGCTGTGGTT 120
Qy 41 IleThrIleAspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThr 60
Db 121 ATTACCATTTGATAGAGCGGTATGGCATTGCCAATCGTATCCGCAAAATGCCCATACC 180
Qy 61 ThrLysValIleTyrGluGluGlnIleGlnGluAlaThrGlySerArgGlnLeuAla 80
Db 181 ACCAAAGTTATTTATGAGAGCAAAATTCAGAGCAAGCAACAGGCTCTCGACAGCTTGC 240
Qy 81 AspValMetAlaGlnLeuIleProSerLeuGlyValSerSerGlyThrThrSerAnPhe 100
Db 241 GATGTGATGGCAGCTCATTTCCAAAGTTTGGGGGTAGTAGTGCACTACCAAGTAACTTT 300
Qy 101 GlyGlnThrMetHisGlyArgGlnValGlnPheLeuLeuAsnGlyValProLeuThrGly 120
Db 301 GGGCAACCATTCACCGTCTCGTCAAGTGCAATTTTGTTAATGGCGTCTTGACAGCT 360
Qy 121 SerArgAspIleSerArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGlu 140
Db 361 TCGCGAGACATCTTAGACAGCTTAATAGTATCAATCCCAATCAAGTGGCTAGAAATTGAA 420
Qy 141 ValLeuSerGlyValAlaThrSerIleTyrGlySerGlyValAlaThrGlyValLeuIleAsn 160
Db 421 GTTTTATCAGGAGCAACCAAGTATTTATGGGTCTGGAGCAACAGCGGTTGATTAAATATC 480
Qy 161 ValThrLysSerAspLeuGluGluGlnPheGlnThrArgIleGlyValHisGlySer 180
Db 481 GTTACTAAGTCTGATTTGGAGAGGAGCAATTTGAAACCGCATCCGTTACATGCTAGT 540
Qy 181 LysLeuSerSerGluGlyIleGlyTyrGlnValGlyGlnSerValAlaGlyValSerGlu 200
Db 541 AAATTATCCAGTCAAGGTATCGGTTATCAGGTAGGTAGGTAGGTAGGTAGGTAGGTAGG 600
Qy 201 AsnGlyAsnValLeuAlaArgLeuAspValAspTyrArgThrThrGlyGlyAlaPheAsp 220

```

```

Db 601 AATGGTAATGTCCTTGCACGACTTGATGTCGACTATCGCACCAAGAGGGGCATTTGAT 660
Qy 221 AlaAsnGlyLysArgIleAlaProGluProAlaGlnThrAspLysGlnAspSerLysSer 240
Db 661 GCTAACGGTAAACGCATCGCTCTCGAGCTGCCCCAAACTGATAAGCAAGACAGCAAAAGC 720
Qy 241 LeuSerValAsnThrAsnValAspTrpGlnLeuAspLysGlnAsnIleAsnLeuAla 260
Db 721 CTAAGTGTCAATACAAATGTTGATTGGCACTTCGACCAAGCAAAATATCAATCTGCGCA 780
Qy 261 LeuThrHisTyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeu 280
Db 781 TTGACGCAATTAACGACAAACAAGATACCGATTATGCACCTGATTATGGTAATCGCCTT 840
Qy 281 AlaValLeuPheGlyGluLysProSerLeuAsnAlaIleLysGlyLeuSerLeuSerGlu 300
Db 841 CGCGTGTGTTTGGAGAAAGCCCTTCAATTAATGCCATCAAGGCTTATCATTTATCAGAA 900
Qy 301 GlnProLysThrThrLysSerThrPheAsnIleAsnTyrHisAspAspLeuTrpGly 320
Db 901 CAGCCAAAACCCACCAAAAGCACCTTTAATATCAACTATCATCATGATGATTTGTGGGGT 960
Qy 321 AsnThrIleAsnThrAsnAlaTyrTyrArgArgGluLysGlyArgPheTyrProPheVal 340
Db 961 AACACCATCAATACCAATGCTTATTATCGCAGAGAAAGCAGATTTTATCCCTTTGTT 1020
Qy 341 AlaProPheSerIleAlaLysAlaLeuProIleLeuGlnSerMetAsnLeuProSerAla 360
Db 1021 GCCCGTTTTTCGATCGCCAAAGCCCTGCTATTTTACAAAGCATGAATTTGCCATCAGCC 1080
Qy 361 ThrLeuAspAlaTyrThrLysAlaProGlnAlaArgAlaTyrGlyValLeuGlnSerGlu 380
Db 1081 ACTTTGGATGCTTATACCAAGGCTCCCAAGCTCGCGCTATGGGGTGTACAAATCCGAA 1140
Qy 381 SerLysAlaGluValLeuGlyArgValProAsnLeuAsnLysProLysArgAlaLeuPhe 400
Db 1141 TCTAAGGCAGAGGTACTAGGCGGTGCTCCCTAATTTGAATAAGCCCAAAAGAGCCCTATT 1200
RESULT 2
AX027664      AX027664      1203 bp      DNA      linear      PAT 16-SEP-2000
LOCUS
DEFINITION    Sequence 1 from Patent WO0043519.
ACCESSION    AX027664
VERSION      AX027664.1      GI:10188544
KEYWORDS
SOURCE      Neisseria meningitidis.
ORGANISM    Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
            Neisseria.
REFERENCE    1 (bases 1 to 1203)
AUTHORS      Ruelle,J.L.
TITLE        Novel compounds
JOURNAL      Patent: WO 0043519-A 1 27-JUL-2000;
            RUELE JEAN LOUIS (BE) ; SMITHKLINE BEECHAM BIOLOG (BE)
FEATURES
source
BASE COUNT    359 a 261 c 263 g 320 t
ORIGIN
Alignment Scores:
Pred. No.:    2,94e-142      Length:    1203
Score:        2047.00      Matches:    400
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 100.00%      Indels: 0
Gaps: 0
DB: 1
US-09-889-746-2 (1-400) x AX027664 (1-1203)
Qy 1 MetArgHisSerHisTyrPheGlnTrpLeuSerLeuProLeuLeuSerValAlaValThr 20

```

```

Db      1 ATGGCGTATCCCATATTTTCATATGTTATCTTGTACTTACGATGGGACGTAACCT
Qy      21 GINGINLeuTyraGInProAnGInSerLeuProThyAIGInLeuGInProVal1 40
Db      61 CAGCAGATTGACCTCAACCAATGAGTCATTCACCAAGGATTAAATTAGAGCTGTGGT 120
Qy      41 1LeThr1eAsp1ySerG1yMe1a1eAsnA1g11eThrG1mEcPro1sThr 60
Db      121 ATTACATTGATAGAGCGGTATGCGACTTGCCTATGATTCAGCAAGGCCCATACC 180
Qy      61 Thr1yVal11eTyrg1uG1n11eG1nG1n1a1ThrG1ySerA1g1n1eUa1a 80
Db      181 ACCAAGTATTATATAGAGCAAAATTCAGAGCAACCAAGCAAGCTCTCGACAGCTTGGC 240
Qy      81 AsnValMe1a1GIn1e1eProSerLeuG1yVal1SerSerG1yThrThrSerAnPhe 100
Db      241 GATGTATGCGACAGCTCATTCCAAGTTGGGGTAAAGTGGCACTACCACTACTT 300
Qy      101 G1yG1nThrMe1a1G1yATG1nVal1G1nPhel1eUeAsnG1yVal1ProLeuThrG1y 120
Db      301 GGGCAACCATGACCGTCTCTCAAGTCAATTTTGTAAATGGCGCTTTCAGACAGT 360
Qy      121 SerA1gA1p11eSerA1g1n1eUeAsnSer11eAsnProAnG1nVal1A1a1g11eG1n 140
Db      361 TCGCGAGACATCTAGACAGCTTAATAGTATCAATCCCAATCAAGTGGCTAGAA 420
Qy      141 Val1eUeSerG1yA1ThrSer11eTyrg1ySerG1yA1ThrG1yG1yLeu11eAsn11e 160
Db      421 GTTTTACAGAGCAACCAATATTATTAAGGATCGAGCAACAGCGGTTGATTAATC 480
Qy      161 ValThr1ySerA1p1eUeG1nG1nG1nPhel1eUThrA1g1yVal1H1eG1ySer 180
Db      481 GTTACTAAGCTGATTTGGAAGAGCAATTCAAACCCGCACTGATCAATGAGT 540
Qy      181 LyeUeSerSerG1yG1y11eG1yTyrg1nVal1G1yInSerVal1A1g1yVal1SerG1n 200
Db      541 AAATTATCCAGTAAAGGATATCGGTTATCAGGTAGTCAAGTGTAGCAGGTGTACGAA 600
Qy      201 AsnG1yAsnVal1eUa1a1eUeAsnVal1AspThrA1g1ThrG1yG1yA1a1PheAsp 220
Db      601 AATGGTATATCTCTGACACATGATGATGATGATGATGATGATGATGATGATGAT 660
Qy      221 A1aAsnG1y1yA1g11eA1a1eProG1yProA1a1GInThrA1p1ySerG1nA1p1ySer 240
Db      661 GCTAACGCTAAAGCATCGCTCTGAGCTGCGCAAACTGATAGCAAGACAGCAAAAGC 720
Qy      241 LeuSerVal1AsnThrAsnVal1Asp1rG1n1eUeAsp1yG1nAsn11eAsn1eUa1a 260
Db      721 CTAAAGTCAATACAAATGTTGATGCACTTGACGCAACCAAGCAAAATATCAATCTGCA 780
Qy      261 LeuThr1s1y1rAenA1p1yG1nAspThrA1p1yA1a1PheA1p1yG1yA1a1rG1eU 280
Db      781 TTGACGATTATTAACCAACCAAGATACCAATATCACTCTGATTAATGATGATGATGAT 840
Qy      281 A1aVal1eUePheG1yG1yU1yP1ySerSerLeuAsnA1a11eUeG1yLeuSerLeuSerG1n 300
Db      841 GCGGTGTTGTTGAGAAAGCCCTTCAATTAATGCAATCAAGGCTTATCTTTCGAA 900
Qy      301 G1nPro1yThrThr1ySerThrPheAsn11eAsnTy1sH1eAspA1p1yLeu1rG1y 320
Db      901 CAGCAAAACCAACCAAGACCTTAAATATCAATATCAATGATGATGATGATGATGAT 960
Qy      321 AsnThr11eAsnThrAsnA1a1Ty1yA1a1rG1yU1yG1yA1a1p1yThrP1yProPheVal 340
Db      961 AACCAATCATATACCAATGCTTATATCGAGAGAAAGCAATTTTATCCCTTGT 1020
Qy      341 A1aProPheSer11eA1a1eUeA1a1eUePro11eUeG1nSerMe1a1eUeProSerA1a 360
Db      1021 GCCCGTTTTCATGCGCAAGCCCTGCTATTTTACAAAGCATGAATTTGCCATCAGCC 1080
Qy      361 Thr1eUeAspA1a1Ty1yThr1yA1a1ProG1n1a1a1rG1a1Ty1yVal1eUeG1nSerG1n 380

```

```

Db      1081 ACTTGATGCTTATATACCAAGCTCCACAGACTCGCGCTATGCGGTGTACATCCGA 1140
Qy      381 Ser1yA1a1G1yVal1eUeG1yA1rG1yVal1ProAn1eUeAsn1yPro1yA1a1eUePhe 400
Db      1141 TCTTAAGCAGAGTACTAGGCGGTCTCTTATTTGAATTAACCCAAAGACCCCTATTT 1200

RESULT 3
AC098122
LOCUS
DEFINITION
Rattus norvegicus clone CH230-133H7, *** SEQUENCING IN PROGRESS
*** 54 unordered pieces.
AC098122 162976 bp DNA linear HTG 12-JUL-2002
AC098122.3 GI:21729207
VERSION
HTG, HTGS PHASE1.
KEYWORDS
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 162976)
REFERENCE
Muzny D.M., Adams C., Adio-Oduola B., Alt-osman F.R., Allen C.,
Albrooks S.L., Amarantunga H.C., Are J.R., Ayele M., Banks T.,
Barbaria J., Benton J., Bimage K., Blankenburg K., Bonnin D.,
Bouck J., Bowie S., Brileva M., Brown E., Brown M., Bryant N.P.,
Buhay C., Burch P., Burkett C., Butrell K.L., Byrd N.C.,
Carron T.F., Carter M., Cavazos S.R., Chacko J., Chavez D.,
Chen G., Chen R., Chen Z., Chowdhry I., Christopoulos C.,
Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R.,
Davila M.L., Davis C., Davy-Carroll L., Dederich D.A.,
Delaney K.R., Delgado O., Dem A.U., Ding Y., Dinh H.H.,
Douthwaite K.J., Draper H., Dugan-Rocha S., Durbin K.J.,
Barthart C., Edgar D., Edwards C.C., Elhaj C., Escotte M.,
Falls T., Ferraguto D., Flagg N., Ford J., Foster P., Frantz P.,
Gabisi A., Gao J., Garcia A., Garner T., Garza N., Gill R.,
Correll J.H., Guevara W., Gunaratne P., Hale S., Hamilton K.,
Harris C., Harris K., Hart M., Haylak P., Hawes A., Hernandez J.,
Hernandez O., Hodgson A., Hognes M., Holloway C., Hollins B.,
Hornel F., Howard S., Huber J., Huiyk S., Hume D., Jackson L.B.,
Jacobson B., Jia Y., Johnson R., Jolivet S., Joudah S.,
Karlisson E., Kelly S., Khan U., King L., Korvah J., Kovar C.,
Kratovic J., Kureshi A., Landry N., Leal B., Lewis L.C., Lewis L.,
Li J., Li Z., Lichtarge O., Lieu C., Liu J., Liu W., Loulsesed H.,
Lozado R.J., Lu X., Lucier A., Lucier R., Luna R., Ma J.,
Maheshwari M., Mapua P., Martin R., Martindale A., Martinez B.,
Massey E., Mawhinney B., McLeod M.P., Meador M., Mei G., Metzger M.,
Miner G., Miner Z., Mitchell T., Mohabbat K., Morgan M., Morris S.,
Moser M., Neal D., Newton J., Newton N., Nguyen A., Nguyen N.,
Onguyen N., Nickerson E., Nwokwenkwo S., Ogun M., Okunolu G.,
Oragunye N., Oviedo R., Pace A., Payton B., Peery J., Perez L.,
Peters L., Pickens R., Primus B., Pu L.L., Quiles M., Ren Y.,
Rives M., Rojas A., Rojibokan I., Rolfe M., Ruiz S., Savery G.,
Scherer S., Scott G., Shen H., Shooshbari N., Sison I.,
Sodergren B., Sonalike I., Sparks A., Stanley H., Stone H.,
Sutton A., Svatek A., Tabot P., Tamerisa A., Tamerisa K., Tang H.,
Tansey J., Taylor C., Taylor T., Telford B., Thomas N., Thomas S.,
Uemami K., Vasquez L., Vera V., Villalón D., Vinsom R., Wang Q.,
Wang S., Ward-Moore S., Warren R., Washington C., Watlington S.,
Williams G., Williamson A., Wlezyk R., Wooden S., Worley K.,
Wu C., Wu Y., Wu Y.F., Zhou J., Zorilla S., Nelson D.,
Weinstock G. and Gibbs R.
Direct Submision
Unpublished
2 (bases 1 to 162976)
REFERENCE
Worley K.C.
Direct Submision
Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 162976)
REFERENCE
Worley K.C.
Direct Submision
Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One

```

## COMMENT

Baylor Plaza, Houston, TX 77030, USA  
On Jul 11, 2002 this sequence version replaced gi.117975686.

----- Genome Center  
Center: Baylor College of Medicine  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu

## ----- Project Information

Center project name: GGLO  
Center clone name: CH230-133H7  
----- Summary Statistics

Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 122651 bases at least Q40

Consensus quality: 125497 bases at least Q30

Consensus quality: 127787 bases at least Q20

-----  
\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 54 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 1216: contig of 1216 bp in length

\* 1217 1316: gap of unknown length

\* 1317 2494: contig of 1178 bp in length

\* 2495 2594: gap of unknown length

\* 2595 3705: contig of 1111 bp in length

\* 3706 3805: gap of unknown length

\* 3806 4856: contig of 1051 bp in length

\* 4857 4956: gap of unknown length

\* 4957 6208: contig of 1252 bp in length

\* 6209 6308: gap of unknown length

\* 6309 8040: contig of 1732 bp in length

\* 8041 8140: gap of unknown length

\* 8141 9439: contig of 1299 bp in length

\* 9440 9539: gap of unknown length

\* 9540 11182: contig of 1643 bp in length

\* 11183 11282: gap of unknown length

\* 11283 12672: contig of 1390 bp in length

\* 12673 12772: gap of unknown length

\* 12773 14429: contig of 1657 bp in length

\* 14430 14529: gap of unknown length

\* 14530 16252: contig of 1723 bp in length

\* 16253 16352: gap of unknown length

\* 16353 17490: contig of 1138 bp in length

\* 17491 17590: gap of unknown length

\* 17591 18953: contig of 1363 bp in length

\* 18954 19053: gap of unknown length

\* 19054 20759: contig of 1706 bp in length

\* 20760 20859: gap of unknown length

\* 20860 22827: contig of 1968 bp in length

\* 22828 22927: gap of unknown length

\* 22928 24503: contig of 1576 bp in length

\* 24504 24603: gap of unknown length

\* 24604 26222: contig of 1619 bp in length

\* 26223 26322: gap of unknown length

\* 26323 29204: contig of 2882 bp in length

\* 29205 29304: gap of unknown length

\* 29305 31498: contig of 2194 bp in length

\* 31499 31598: gap of unknown length

\* 31599 33022: contig of 1424 bp in length

\* 33023 33122: gap of unknown length

\* 33123 34372: contig of 1250 bp in length

\* 34373 34472: gap of unknown length

\* 34473 36026: contig of 1554 bp in length

\* 36027 36126: gap of unknown length

\* 36127 38351: contig of 2225 bp in length

\* 38352 38451: gap of unknown length

\* 38452 40637: contig of 2186 bp in length

\* 40638 42474: gap of unknown length

\* 42475 42574: contig of 1737 bp in length

\* 42575 44551: gap of unknown length

\* 44552 44551: contig of 1877 bp in length

\* 44552 46897: gap of unknown length

\* 46897 46997: contig of 2346 bp in length

\* 46998 49833: gap of unknown length

\* 49833 49833: contig of 2836 bp in length

\* 49834 53440: gap of unknown length

\* 53440 53440: contig of 3507 bp in length

\* 53441 55812: gap of unknown length

\* 55812 55912: contig of 2272 bp in length

\* 55913 58981: gap of unknown length

\* 58981 59081: contig of 3069 bp in length

\* 59081 62244: gap of unknown length

\* 62244 62344: contig of 3163 bp in length

\* 62345 66049: gap of unknown length

\* 66049 66149: contig of 3705 bp in length

\* 66150 69566: gap of unknown length

\* 69566 69666: contig of 3417 bp in length

\* 69667 74553: gap of unknown length

\* 74553 74653: contig of 4893 bp in length

\* 74653 79476: gap of unknown length

\* 79476 79576: contig of 4817 bp in length

\* 79577 83169: gap of unknown length

\* 83169 83269: contig of 3593 bp in length

\* 83270 87095: gap of unknown length

\* 87095 87196: contig of 3827 bp in length

\* 87197 90969: gap of unknown length

\* 90969 91069: contig of 3773 bp in length

\* 91070 94679: gap of unknown length

\* 94679 94779: contig of 3610 bp in length

\* 94780 95032: gap of unknown length

\* 95032 99132: contig of 4253 bp in length

\* 99133 103421: gap of unknown length

\* 103421 103421: contig of 4289 bp in length

\* 103422 107479: gap of unknown length

\* 107479 107479: contig of 3958 bp in length

\* 107480 111787: gap of unknown length

\* 111787 111787: contig of 4208 bp in length

\* 111788 111887: gap of unknown length

\* 111888 115242: contig of 3355 bp in length

\* 115243 115342: gap of unknown length

\* 115343 119864: contig of 4522 bp in length

\* 119865 119964: gap of unknown length

\* 119965 124466: contig of 4502 bp in length

\* 124467 124566: gap of unknown length

\* 124567 129116: contig of 4550 bp in length

\* 129117 129216: gap of unknown length

\* 129217 135370: contig of 6154 bp in length

\* 135371 135470: gap of unknown length

\* 135471 141729: contig of 6259 bp in length

\* 141730 141829: gap of unknown length

\* 141830 146723: contig of 4893 bp in length

\* 146723 146822: gap of unknown length

\* 146823 151589: contig of 4767 bp in length

\* 151590 151689: gap of unknown length

\* 151690 157510: contig of 5821 bp in length

\* 157511 157610: gap of unknown length

\* 157611 162976: contig of 5366 bp in length.

## Alignment Scores:

Pred. No.: 1.15e-36 Length: 162976  
Score: 652.00 Matches: 144  
Percent Similarity: 61.32% Conservative: 70  
Best Local Similarity: 41.26% Mismatches: 117  
Query Match: 31.85% Indels: 18  
DB: 2 Gaps: 9

US-09-889-746-2 (1-400) x AC098122 (1-162976)



QY 1 MetArgHisSerHisSerIlePheGlnIlePheLeuSerLeuProLeuLeuSerValAlaValThr 20  
 DB 1407 ATGCGGCAATACC-----GCAGCAACCTTTCCTGCGCGATGCTGGCGGATTCAGT 1460  
 QY 21 GlnGlnLeuThrLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 39  
 DB 1461 GGCACG 1520  
 QY 40 ValIleThrIleAspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHis 59  
 DB 1521 AGGTGACCG 1568  
 QY 60 ThrThrLysValIleTyrGlnGlnGlnIleGlnGlnGlnAlaThrGlySerArgGlnLeu 79  
 DB 1569 ACCGTATATGCGATGCG 1628  
 QY 80 AlaAspValMetAlaGlnLeuIleProSerLeuGlyValSerSerGlyThrThrSerAsn 99  
 DB 1629 GCCGACATCTCG 1688  
 QY 100 PheGlyGlnThrMetHisGlyArgGlnValAlaPheLeuLeuAsnGlyValProLeuThr 119  
 DB 1689 TACGGATATACCATGCG 1748  
 QY 120 GlySerArgAspIleSerArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArgIle 139  
 DB 1749 GCGCGACGCGATGCG 1808  
 QY 140 GlnValLeuSerIleValAlaThrSerIleTyrGlySerGlyValAlaThrGlyLeuIleAsn 159  
 DB 1809 GAGGTATATCTCG 1868  
 QY 160 IleValThrLysSerAspLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 178  
 DB 1869 ATCATCATCTGACAGCG 1928  
 QY 179 Gly---SerLysLeuSerSerGlnGlyIleGlyTyrGlnValGlyGln-----SerVal 195  
 DB 1929 AGCCCTCGCATCTGTCACAGTACACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1988  
 QY 196 AlaGlyValSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 215  
 DB 1989 CCGCGCGCATCATGTCAGTGC-----GCCGTCGCGCGCGCGCGCGCGCGCGCGCGCG 2036  
 QY 216 GlyGlyAlaPheAspAlaAsnGlyLysArgIleAlaAlaProGlnProAlaGlnThrAspLys 235  
 DB 2037 GCGCGCATCTGCGATGCGATGCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2096  
 QY 236 GlnAspSerLysSerLysSerLysValAsnThrAsnValAspTyrGlnLeuAspAspLysGln 255  
 DB 2097 CAGGACACCGGACCG 2156  
 QY 256 AsnIleAsnLeuAlaLeuThrHisTyrAsnAspLysGlnAspThrAspTyrAlaProAsp 275  
 DB 2157 CACTTGAAGCGTGGCG 2216  
 QY 276 TyrGlyAlaAsnArgLeuAlaValLeuPheGlyGln-----LysProSerLeuAsnAlaIle 293  
 DB 2217 TATGCG 2276  
 QY 294 LysGlyLeuSerLeuSerGlnGlnProLysThrThrLysSerThrPheAsnIleAsnTyr 313  
 DB 2277 CCGGCG 2336  
 QY 314 HisHisAspAsp---LeuTyrGlyAsnThrIleAsnThrAsnAlaTyrTyrArgArgGln 332  
 DB 2337 CGGACCGCGCATGCTACTGCG 2396  
 QY 333 LysGlyArgPheTyrProPheValAla 341  
 DB 2397 CAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2423  
 RESULT 4

AB004565  
 LOCUS AB004565 11579 bp DNA linear BCT 30-AUG-2000  
 DEFINITION Pseudomonas aeruginosa PA01, section 126 of 529 of the complete genome.  
 ACCESSION AB004565 AB004091  
 VERSION AB004565.1 GI:9947294  
 KEYWORDS  
 SOURCE Pseudomonas aeruginosa.  
 ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.  
 REFERENCE 1 (bases 1 to 11579)  
 AUTHORS Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J., Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Lardig,K., Lim,R., Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.  
 TITLE Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
 JOURNAL Nature 406 (6799), 959-964 (2000)  
 MEDLINE 20437337  
 PUBMED 10984043  
 REFERENCE 2 (bases 1 to 11579)  
 AUTHORS Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J., Lagrou,M., Garber,R.L., Goltry,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Lardig,K., Lim,R.M., Smith,K.A., Spencer,D.H., Wong,G.K., S., Wu,Z., Paulsen,I.T., Reizer,J., Salier,M.H., Hancock,R.E.W., Lory,S. and Olson,M.V.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA  
 FEATURES  
 source location/Qualifiers  
 1..11579  
 /organism="Pseudomonas aeruginosa"  
 /strain="PA01"  
 /db\_xref="taxon:287"  
 complement(115..528)  
 /gene="PA1353"  
 complement(115..528)  
 /gene="PA1353"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAG04742.1"  
 /db\_xref="GI:9947295"  
 /translation="MQIVPYLIPNGNCREAFSCYHQHIGTLAHLPGDSPECGDIPADMKIKIHARLVGSGFALMASDNDHPAYPGIGKCSISLVNDSVAAEALRPLNALAGGSVQPLPPTFMAASFQMTDFRFGVAMVNCEDR"  
 complement(565..990)  
 /gene="PA1354"  
 complement(565..990)  
 /gene="PA1354"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAG04743.1"  
 /db\_xref="GI:9947296"  
 /translation="MRPMVIVKATQASRAGEMPSEELLAAMGRYVEELVKAGVLAGLQPSAKGARVRFSGNARTVIDGPAETREILAGPMFQVSLSEALIEWYKRCRPPFDGSELRIOVEAREDFGEFTBELREBERLRALREKRN"  
 1137..1517  
 /gene="PA1355"  
 1137..1517  
 /gene="PA1355"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical protein"  
 /protein\_id="AAG04744.1"  
 /db\_xref="GI:9947297"  
 /translation="WADERRCGVCRLGGRPQWPWPAGLILHCSRAIPASLCLMPRIIVA

```
gene
CDS
PAIORHVAMPALDVGAATVAAALQAAFAREPRRLGCLYLLDDOGSLRRHVALFVDCVQVR
DRRLGDPLODDSEIYVQALSGG"
/genes="PA1361"
/codon_start=1
/transl_table=11
/product="probable transporter"
/protein_id="AAG04750.1"
/db_xref="GI:9947303"
/translation="MNSPALPLSRGLRIRAELEKELITLAAPINIAQLATTAMGFVDVAV
MAGRASPHDLAAVALGNSIWI PMFLMTGTLLATTAQVQRHAGAGDQPGTGPLVRQAL
WLLALIGPLSGAWLWMLSEPIGLMKVRPELIGPSLLYKLGIALGFPAAALYHVRQY
TNGLGRTRPSMVLGIGLLINI PINVALIYGHFGMPKMGPGGCGWATGSVWMPFLCM
LPWNKASLYRASQLPSRWMPDRATIGLVAVLPIGIAVFAESSIFSVIALLIGEL
DENVAHQIALNFSALVFMIPYSLGMANTVRVGHNLGAGLPRDPAAGVGNMAALG
YACVASLMLLREQIAAMSPDPAVIAAASLIVFSALFQSDALQVTAAGALRGYQ
DTRVTMTITFYAYGIGLPGVSGISLGTDFMQEPTPRGLWQGLVVGLTGAAILCIRL
ARGARFIRQHERLQREDAEASVLGR"
7018..7341
/genes="PA1362"
/genes="PA1362"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG04751.1"
/db_xref="GI:9947304"
/translations="MAKADDTSNPAITAQSPFTPIPLRRTSQCNQHPCATLYLDS9A
ASDLDFTARRRLESAITGLSVLDDLETEPGEISCRITGSLLSLSDALSLEAF5Y
ARRHR"
7354..8094
/genes="PA1363"
7354..8094
/genes="PA1363"
/codon_start=1
/transl_table=11
/product="probable sigma-70 factor, ECP subfamily"
/protein_id="AAG04752.1"
/db_xref="GI:9947305"
/translations="MPPKAGLSSGAKRGSLSALAKQLRGRKATCPLRTTQAIPO
DRSFICDCFSFTLDPHSSPSPGFCVFPDPAKPTLLSLVRYELVDHVRREFG
DRGWARVVDVQVLLERDEKEDVHTPLALLRKISHDAASVRSRRRTATWVEAVA
ELPEVACMAATPMGRHEAREFOALVDIAIATLPRCQAVFMHKKIHEVPPQAVADRLG
ISVKTVEKHLRLGLAACRERLGRGEARP"
8091..8933
/genes="PA1364"
8091..8933
/genes="PA1364"
/codon_start=1
/transl_table=11
/product="probable transmembrane sensor"
/protein_id="AAG04753.1"
/db_xref="GI:9947306"
/translations="MTQPSKRNQGVPLDDALERHRDALKOLFPLPPSRPKSRKASKAA
SVALLLALGALAWLDPAYKRERFATAIGERDVTLDGSLDLSGSOIEVSWHPL
GRVALLPAGOLFEVSPARVRVSVGSTDIVLGRFNVRLDDRVVTLERGRVE
VGT5ATAQAPVULLPGQVDAIGQLMPLAKVDAATAMAKDDRLVFRTPPLGRAVAL"
gene
CDS
US-09-889-746-2 (1-400) x AB004565 (1-11579)
Alignment Scores:
Pred. No.: 6.88e-29 Length: 11579
Score: 527.00 Matches: 137
Percent Similarity: 51.12% Conservative: 68
Best Local Similarity: 34.16% Mismatches: 144
Query Match: 25.74% Indels: 52
DB: 1 Gaps: 9
Qy 2 ArgHisSerHisTyPheGlnTrpLeuSerLeuProLeuLeuSerValalaval---Thr 20
Db 9290 AGGGACGAACATACCTTC-----ATCGTCGTTCGGCCGACGAGCCGCGCCGCT 9343
Qy 21 GlnGlnLeuTyAlaGlnProAsnGlnSerLeuProThrValGlnLeuGlnProValVal 40
Db 9344 ACCGAAGCCCGACGCAACCG-----CTGGACATCGACGACATGGA 9385
```

QY 41 ILeThrIleAspIysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHleThr 60  
 Db 9386 ATCACCCTCGCGCACACGACGACCTGCTCCGCGCACGCGCGAG----- 9433  
 QY 61 ThrIleValIleIleTyrGlnGlnGlnIleGlnGlnAlaIleThrIleSerArgIleLeuAla 80  
 Db 9434 TCACCGCTTCATCAGCATGCCCACTGGAGAGACTGGCCAGGAGCTCCGACAGCTGGCC 9493  
 QY 81 AspValMetAlaGlnLeuIleProSerLeuGlyValSerSerGlyThrThrSerAsnPhe 100  
 Db 9494 ACCGTGCTGGCCACAGGAGGTCGCGGAGATCTCGATTCCAGCCGCCACCATCACCGAATAC 9553  
 QY 101 GlyIleThrMetHieGlyArgGlnValGlnPheLeuLeuAsnGlyValProLeuThrGly 120  
 Db 9554 GGCACAGACCTCGCGCGGCGCAGCATCTGATGTGTGATGCGCGCGTCCGCGTGAACACC 9613  
 QY 121 SerArgAspIleSerArgIleLeuLeuSerIleAsnProAsnGlnValAlaIleGlu 140  
 Db 9614 AACCGGACTCTCGCGCAACCTGGCCACATCGACCGCGCTGATCGACGCGATCGAG 9673  
 QY 141 ValLeuSerGlyAlaThrSerIleTyrGlySerGlyValThrGlyIleLeuIleAsnIle 160  
 Db 9674 GTATCTCGTGGCAGCAGCGCATCTACGAGAGCGCCACCGCGGATCATCTCCATC 9733  
 QY 161 ValThrIleSerAspLeuGlnGlnGlnPheGlnThrArgIleGlyValIleGly--- 179  
 Db 9734 ACCACCGCTCGCGCGCGCGGAGAACCGCGGAAACCGCGCTCAGCGGCACCTCGCG 9793  
 QY 180 ---SerIleLeuSerSerGlyGlyIleGlyTyrGlnValGlnIleValAlaGlyVal 198  
 Db 9794 CTGACCGCTGGGACGCGCATGCTCGCGCGCATGCTCCAGCATCTTCGCGCGCG--- 9850  
 QY 199 SerGlyAsnGlyAsnValIleAlaArgLeuAspValAspTyr-----ArgThr 214  
 Db 9851 -----TCCCTGGGCGCTCGACTATTCTTCGACTTCGCGACCCCGCAC 9895  
 QY 215 ThrGlyGlyAlaPheAspAlaAsnGlyIleArgIleAlaIleProGlnProAlaGlnThrAsp 234  
 Db 9896 GTCGGCGCTTCTTACAGACGCGCATGCGCAGCGCATGCGCGGACCCGACCGAGCGGAC 9955  
 QY 235 LysGlnAspSerIleSerLeuSerValAsnThrAsnValAspTyrGlnLeuAspAspLys 254  
 Db 9956 CTGTTCGACTCGAACGCTTCAACATCGCGGCGAAGCTTCGCGCTCGCGCATCGACGAGAAC 10015  
 QY 255 GlnAsnIleAsnLeuAlaLeuThrHleTyrAsnAspLysGlnAspThrAspTyrAlaPro 274  
 Db 10016 CACCGGCTGACGCTCGCTCGCTCAGCATACGAGCGCGCGACGACGACTACGCGCAC 10075  
 QY 275 AspTyrGlyAsnArgLeuAlaValIleuPheGlyGlyLysPheSerLeuAsnAlaIleLys 294  
 Db 10076 GAC-----CCGAGGGTGGCGCAGGCTCGCGCGCGCTCGCGCGCCACCGCATCAA 10129  
 QY 295 GlyLeuSerLeuSerGlnGlnProLysThrThrIleSerThrPheAsnIleAsnTyrHle 314  
 Db 10130 GGGCTGAGCTGAGCGAGCAACCGCATCCGCAACACCTCGCGAAGCTCGAGTACGAG 10189  
 QY 315 HisAspAspLeuTyrGlyAsnThrIleAsnThrAsnAlaTyrTyrArgArgGlyLysGly 334  
 Db 10190 AACCTGACATCTCGCGACGCGGCTCTCGCGCAGCTTACTACGCGAGCTATTTCACC 10249  
 QY 335 ArgPheTyrProPheValAlaProPheSerIleAlaIleValAlaLeuProIleLeuGlnSer 354  
 Db 10250 CGCTTACATCTCGTTCCAGCGC----- 10270  
 QY 355 MetAsnLeuProSerAlaThrLeuAspAlaTyrThrIleValAlaProGlnAlaArgAlaTyr 374  
 Db 10271 -----CGCCCGCTCTCCACCGCGCGCGC 10294  
 QY 375 GlyValLeuGlnSerGlnSerLysAlaGlnValLeuGlyArgValProLeuLeuAsnLys 394  
 Db 10295 AATGTGACCATCATCATGCAAGAACGAGGAGTGTTCGCGAGCGCGCTGACCTGCGCAGC 10354

QY 395 Pro 395  
 Db 10355 CCG 10357  
 RESULT 5  
 AEO04881  
 LOCUS  
 DEFINITION  
 Pseudomonas aeruginosa PA01, section 442 of 529 of the complete genome.  
 AEO04881  
 GI:9950926  
 AB004881.1  
 GI:9950926  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Pseudomonas aeruginosa.  
 Pseudomonas aeruginosa  
 Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
 Pseudomonas.  
 1 (bases 1 to 12421)  
 Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S.L., Hutnagle,W.O., Kowalik,D.J., Lagrou,M., Garber,R.L., Goltzy,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Lardig,K., Lim,R., Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T., Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
 Nature 406 (6799), 959-964 (2000)  
 2 (bases 1 to 12421)  
 Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S.L., Hutnagle,W.O., Kowalik,D.J., Lagrou,M., Garber,R.L., Goltzy,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Lardig,K., Lim,R.M., Smith,K.A., Spencer,D.H., Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saiter,M.H., Hancock,R.E.W., Lory,S. and Olson,M.V.  
 Direct Submission  
 Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA  
 Location/Qualifiers  
 1..12421  
 /organism="Pseudomonas aeruginosa"  
 /strain="PA01"  
 /db\_xref="taxon:287"  
 89..703  
 /gene="PA4671"  
 89..703  
 /gene="PA4671"  
 /codon\_start=1  
 /transl\_table=11  
 /product="probable ribosomal protein L25"  
 /protein\_id="AAG08058.1"  
 /db\_xref="GI:9950927"  
 /translation="MWDFILNVAQVRSDLGKASRRRLRRNAGLVPAVVYGGDEKPEOST  
 LELRIKILLENLAEAFVIALNVGAGKSTVLIKALRRHPAKGFVMDPFLVADHK  
 LTAHVLPILNEVAFVGVKAGGKSTSHTSIEVSCLPKRDPEFLVDMAKVELQIV  
 HSLDKAPKGVLEVLQALNDLAVANIIASRVKEGSEBGALE"  
 745..1329  
 /gene="PA4672"  
 745..1329  
 /gene="PA4672"  
 /codon\_start=1  
 /transl\_table=11  
 /product="peptidyl-L-trna hydrolase"  
 /protein\_id="AAG08059.1"  
 /db\_xref="GI:9950928"  
 /translation="MTAVQLYGLGNPPEYDQTHNAGALFVERLHAQGVSLVADR  
 KYFGLVGRSHQGVRLIPPTVYNRSGSVPAALAGFRLAPDAIVLADHLDMPG  
 VAKTKTGGHGHNGLRDIIQIQNSPHRLRLGICHPGHSLSVGYVLGAPRSEQ  
 ELIDSTIPALGVLEPMLAGDWTTRMQLHSOKA"  
 1370..2470  
 /gene="PA4673"  
 1370..2470  
 CDS

```

/gene="PA4673"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG08060.1"
/db_xref="GI:9950929"
/translation="MGFNCGLVGLPNVCKSTLFLNALTKSGTGAENPFPTCTIEPNNGIV
PMPDARLNALAEIVKPERVLPTTFEFVDIAGLVAGSKGEGKLNFLANRETDAIAH
VVRPEDNVSDVDPKRDIEIIDLELIFADLSCQKOLQKVARNAKGGDKREALA
KALLEKLIHPHTEGKAPSLKLNGLGEEKLVRSFHLTSPKVMYANVAEDGFENN
PHLDVKAIAEEGAVVVPVCNKTEABTAELEDGEEKDMPLESLGLEPGLNRVIRAG
YGLNLQTYFTAGVKEVRANTVRVGATAPQAQGVHTDFEKGFTIRAEVAVDDFIQK
GEQAKAGKRWLEGGKDYIVKGDGVMMHFRNV"
2650..2726
/product="crna-met"
/complement(2877..3182)
/gene="PA4674"
/complement(2877..3182)
/gene="PA4674"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG08061.1"
/db_xref="GI:9950930"
/translation="MATNGMRPIHPGEITLDEFLMEFDISPAALARALKVSGAPTVDNI
VREQGISADNAIRLGRYFDTSAQFMNLQSEYSLATAYAAANGKQIEHEIBPLLAHG"
3801..6029
/gene="PA4675"
3801..6029
/gene="PA4675"
/codon_start=1
/transl_table=11
/product="probable TonB-dependent receptor"
/protein_id="AAG08062.1"
/db_xref="GI:9950931"
/translation="NPRSIPLRPAPLALSLSFASAPALAAPVEQMVVIGSRAP
TSISELPTGVVIEREQDQQTQAGVPKRALQGLIPGLDQSGRTNNGNLGRSV
LVMTDGLSNRSQRSDISDPNIEIRIEMSGASAVYGGGATGGLIINIVKKGVG
GDTFRTELGRSQSHEDHLRAAQISGNLDFNGLRAIAYOKNGAAYDGGSDQV
LTDITQTDLOVNSVDLMSGLGFTFANGHSILDGLQYVDSYDGDGLDGLNFDALR
GRAPYSIKGGVDLREPESKRHOFNATYHAPVGLHDLVLOAYYRNEKMAFNPFPIR
YSTGAINGYTSYASQDDTDYVGMKIALVKTWERASLYTGVDLDEKFTSDOMLEN
LPLAASGLVASEQAKGRYPDIDTDSRAFQSGSKATDIDLISAGVRQSKSTDV
SDVAAANQQIILANGLTADAAPGGSXDYDNLVNGAIYKLNMQQVWANYSEGFE
LPDPAKYIGFGRYGAADNGHYPLLGQVSNDSPLDGKTKQVELGWRHTDGLDQV
AAFYSWSDKSTKYDSKTLAVLQDDTKRNVGLEQATYWLDDHWQVNGLAISQEK
VDGRLKQDVTSPSKAGAFVGMKDDORSRLQGVRTFNLDNDEPGNKIDGVALFDLL
GTQALPVGTLAGIQLLDKDYTTWGGRAQVYVGLAPAGLFDYKGRGYSLTYSV
E"
/complement(6099..6746)
/gene="PA4676"
/complement(6099..6746)
/gene="PA4676"
/codon_start=1
/transl_table=11
/product="probable carbonic anhydrase"
/protein_id="AAG08063.1"
/db_xref="GI:9950932"
/translation="MSDLQQLFENNVRRAEALKORDPDPFAKLARQQTPEYLIWGCS
ARVPANEIVGMLPGLFVHRVANVYVHTDNLCLSVIQFADVILVKXHLIVTGHYGG
GVRASLNDQGLIDGWLRSIDLAYEYREHLEQLPTEERVDRLCELNVIQQVANVS
HTSIVQNAWHRRGQSLSVGCIYGIKGLWKNLNTVSGLDQLPPOYRLSPILGGCC"
/complement(6808..8046)
/gene="PA4677"
/complement(6808..8046)
/gene="PA4677"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG08064.1"
/db_xref="GI:9950933"
/translation="MCDTLILRHOGQTVFAKNSDREPAEPORLLRLPAVRNDPEARLR
TYYLDIPQARHALVISOQSWIWAEMGVNEHVAIGNAEVFTKLTRQRTALLGMD

```

```

LLRLGRCASAREALEVITGLLQRYCGGPGACRYDKRIRYDNSFLIADPGEAWVLET
AGNLAAKVERWAIENALTGLGHEFDLCSDLEDQARRQCQWGDGFHFARVFDTRL
LPWVGGAHRCRINQRFSLPGEPMHALFAALRHGPRGDFRHRNNQVCMHAGS
FMRPSQTTASLVARLDGDLAATGTSAPCLGLFQPLSPDPAAGAAVLQSPREAVOE
SPWRFVAVHRRALTDRAPRRLRSGRDMLERELFADLEQDPWRLAEADAAAHACHWH
DOVAMOPLAQQRWRWSHAAL"
/complement(8245..8697)
/gene="rim1"
/notes="PA4678"
/complement(8245..8697)
/gene="rim1"
/codon_start=1
/transl_table=11
/product="peptide n-acetyltransferase Rim1"
/protein_id="AAG08065.1"
/db_xref="GI:9950934"
/translation="MSDAISPRHMTADLDAVLKIEYAFSPHWTRGIFSEGLKNYHC
WLMFEGSOQVGHGVIQIILLDEAHLNITVKPESQGLGLRLLEHLHLMREAAGHQAREC
FLEVRASNGAAYRLYERYGNEIGRRRDYYPVGGREDALVMACTLFD"
/complement(8694..9395)
/gene="PA4679"
/complement(8694..9395)
/gene="PA4679"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG08066.1"
/db_xref="GI:9950935"
/translation="MPVRAARAADPAQARYPOAEAPASAPSAVPAVEGRGIPISLP
KPGAPAAKAEPAVPEEAPESPAPQAPPRFALQLQAGSTLLVELPTGEPASR
DPGYLLLRDLRAAGLSDPSRLIGEPVWFLPALLAGNLDQGPQAALEFVQSFVAARMBE
SERSRCLWLVGLPAIRPAGEDEGSLFRELQVDGLGATWAVPGLEALMESFALKGELW
RAMRRVQRVRSSEIQ"
/complement(9655..10191)
/gene="PA4680"
/complement(9655..10191)
/gene="PA4680"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG08067.1"
/db_xref="GI:9950936"
/translation="MAQTHPLAGGIGILAALLFSASAGAEIERSLGDSDLDVVGQ
AGISASLNLENPSQTRCAGCGCARLAQPGSGNNYLVIDDISGRFSFDEVTLDV
VTIKDGGGGAAPNRPVLVGLVKDTSFQDLSTFIAGANKARGTDPGLVQTNLVSARI
DGTLLKQGNVNIPTVAPR"
/complement(10222..11241)
/gene="PA4681"
/complement(10222..11241)
/gene="PA4681"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG08068.1"
/db_xref="GI:9950937"
/translation="MSRACLLAALLAAPPFAHAEMTALDDEQLSDVQAGIGFVLDSAM
VDGNKATIAINGVTNATQGNVPTVKELVYAATGSKNGSTLAPVTILGRNLPFKLNL
AKGSLVTLQDDGRVSTLPTADVIELAPPSLVSGAGGQPCIAAGYAGAGNCSRSAS
EKVDAGIRFPDGVFTSTRDVLNLDLSLAMDMSYLRFWGANARSQLVEIRLNLVAKT
FEINSCGAGSPSCVSAEAQAARTVYLTINAVASVALGYKSKQPLLPSTADGNFTFELP

```

## Alignment Scores:

Pred. No.:	4.07e-27	Length:	12421
Score:	503.50	Matches:	132
Percent Similarity:	49.53%	Conservative:	80
Best Local Similarity:	30.84%	Mismatches:	171
Query Match:	24.60%	Indels:	45
DB:	1	Gaps:	10

US-09-889-746-2 (1-400) x AE004881 (1-12421)

Qy 11 SerLeuProLeuSerValAlaValThrGlnGlnLeuTyrAlaGlnProAenGluSer 30

||||:|||||

Dh	3810	TCCATCCCTTTGAGCGCGGCTCCGCGTGGAGCTTTCCTGTCCCTGTCCCTGTCCGCTTCGCTTTC	3869
Qy	31	LeuProthrValIgluLeuGluProValIleThrIleAspIysSerGlyMetAlaLeu	50
Dh	3870	GCTCTGCTGGCTGCTGCCATCCGGTGGACAGACAGATGGTGGTATGGCTCGCGCGC	3929
Qy	51	AlaAsnArgIleThrGlnMetProHisThrThrIleValIleGlyGluGlnIleGln	70
Dh	3930	CCGACAGCATCACAGGATGGTCCCGGCACGGTCTGGGTATGAGCGCGACACTGGAC	3989
Qy	71	GluGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnIleuIleProSerLeu	90
Dh	3990	CAGCAGACCCAGCGCGCGCGTCCGCTGTAAGAGAGCGCTTGGACATGATCCCGGCGTG	4049
Qy	91	GlyValSerSerGlyThrThrSerAsnPheGlyGlnThrMetHisGlyArgGlnValGln	110
Dh	4050	GATATCGCTCTCCAGGGTCTGTATCCCAACAGCGCGCAACCTCGCGGACGCGTATCG	4109
Qy	111	PheIleuLeuAsnGlyValProIleuThrGlySerArgAspIleSerArgGlnLeuAsnSer	130
Dh	4110	GTAATGATCGACGGGGGTGGCTGTGAACAGTTCACGGGGATACGGGAGTTCACTCC	4169
Qy	131	IleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyValaThrSerIleGlyIle	150
Dh	4170	ATCGACCATTCACACATCGAGCGCATGAGTATGATCCGCGCGCACGCGGTATACGGC	4229
Qy	151	SerGlyAlaThrGlyGlyLeuIleAsnIleValThrLysSerAspIleu--GluGluGlu	169
Dh	4230	GCGCGCGCACCGCGGGATCATCAATCATGTATGACCAAGAGCGGTGCGCGCGACACT	4289
Qy	170	GlnPheGluThrArgIleGlyValHisGlySerLysLeuSerSerGluGlyIleGlyTyr	189
Dh	4290	CGCTTCAACACCGAGCTGGCGCGCGCGACGCGCTTCAGAGCATAGAGACACACACTCG	4349
Qy	190	GlnValGlyGlnSerValaGlyValSerGluAsnGlyAsnValLeu-----AlaArg	207
Dh	4350	CGCGCGCGCGCATGATCACCGC-----GGCAAGCACTGTTCACGCGCGC	4397
Qy	208	LeuAspValAspTyrArgThrThrGlyValaPheAspAlaAsnGlyLysArgIleAla	227
Dh	4398	CTGGCCATCGCTTACAGAAAGAGCGCGCGCTTACAGCGGACAGTGGCGACAGGTACTG	4457
Qy	228	ProGluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsnThrAsnVal	247
Dh	4458	ACGAGACATCACCCAGACCGACCTGCACTACACAGCGTCTGATGATGGCGAGCTT	4517
Qy	248	AspTrpGlnLeuAspAspLysGlnAsnIleAsnLeuAlaLeuThrHisGlyAsnAspLys	267
Dh	4518	GGCTTCACTTTCGCCAAATGGCACAGCTTGATCTGGCTCGCTCAAGTACTACACCTCGGC	4577
Qy	268	GlnAspThrAspTyrAlaProAspTyrGlyValAsnArgLeuAlaValLeuPheGlyGlyLys	287
Dh	4578	TACGACGGCGACCGCGCGCTCGATCTTGGCGCAACTTGATGCCCTCGCGGACGGGCA	4637
Qy	288	ProSerLeuAsnAlaIleLys---GlyLeuSerLeuSerGluGlnProLysThrThrIys	306
Dh	4638	CCC-----TATTCGATCAAGGCGCGGCTGCACTGACCGCGAGACCGGAGAGCAAGCGC	4691
Qy	307	SerThrPheAsnIleAsnTyrHisPheAspAspLeuTrpGlyAsnThrIleAsnThrAsn	326
Dh	4692	CATCATGTTCAACGCCACTTACACAGCGCGGGAAGTCTGGTCAAGACACTTACCTGAG	4751
Qy	327	AlaTyrTyrArgArgGlyLysGlyArgPheTyrProPhe-----	339
Dh	4752	GCTTACTACCGCAACGAGAGATGGCTTCAATCCTTCCATCAATCCGCTACAGCAAT	4811
Qy	340	-----ValAlaProPheSerIleAlaLys-----	347
Dh	4812	ACGCGCGCATCACTACGCGCACTCTACTACTCGCTCGCGTCCACAGACAGACCGACTAC	4871
Qy	348	-----AlaLeuProIleLeuGlnSerMetAsnLeuProSerAlaThrLeu-----	362
Dh	4872	TACGGATGAATCTGGCCCTGGTGTGAAGACTCGGAGACGCGCAAGCTGACTTACGGGGTG	4931

Oy	363	-----	AApAATyArThr-----	-----	LysAlaProGlnIlaArg	372
Db	4992	GACCTGGACCGGGAATAATTCACTGTCCGACCAAGATGCCTTTCACACTGCCCGCCTGCCGCC				4991
Oy	373	AlATyGlyValLeuGlnSerCluSerLysAlaGluValLeuGlyArgValProAsnLeu				392
Db	4992	GCCAGCGCGGGCTGTGTGTACGTGAAGAAGGCCAAG---CTGGGCGCTCATCCCGGACATC				5048
Oy	393	AenlySProLySArgAlaLeuPhe				400
Db	5049	GATACCAGCACCGCGCTTCTTC				5072
RESULT 6						
LOCUS	AL646078	20350 bp	DNA	linear	BCT 07-DEC-2001	
DEFINITION	Ralstonia solanacearum GM11000 megaplasmid, complete sequence;					
ACCESSION	AL646078	AL646053				
VERSION	AL646078.1	GI:17430778				
KEYWORDS	.					
SOURCE	Ralstonia solanacearum.					
ORGANISM	Bacteria; Proteobacteria; beta subdivision; Ralstonia group;					
REFERENCE	Ralstonia.					
AUTHORS	1 (bases 1 to 203050)					
TITLE	Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,					
JOURNAL	Atiat,M., Billault,A., Brotter,P., Camas,J.C., Catolico,L.,					
REFERENCES	Chandler,M., Choise,N., Claude-Renard,C., Cunne,S., Demange,N.,					
AUTHORS	Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schlex,T.,					
TITLE	Sigulier,P., Thebaud,P., Whalen,M., Wincker,P., Levy,M.,					
JOURNAL	Weissenbach,J. and Boucher,C.A.					
	Genome sequence of the plant pathogen Ralstonia solanacearum					
	unpublished					
	2 (bases 1 to 203050)					
	Boucher,C.A.					
	Direct Submission					
	Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston					
	Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie					
	Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,					
	BP27, 3126 Castanet-Tolosan Cedex, France, Fondation Jean					
	Dausset-CBPH, 27 rue Juliette Dodu, 75010 Paris, France, ILMGM CNRS					
	URVJ, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France, et INRA					
	Laboratoire de Biométrie et Intelligence Artificielle INRA, BP27,					
	F31326 Castanet-Tolosan Cedex, Laboratoire de Génétique Cellulaire					
	INRA, BP27, F31326 Castanet-Tolosan Cedex					
	Christian.Boucher@toulouse.inra.fr					
	http://sequence.toulouse.inra.fr/R.solanacearum.html.					
COMMENT	Location/unaffilers					
FEATURES	1..203050					
SOURCE	/organism="Ralstonia solanacearum"					
	/strain="GM11000"					
	/db_xref="taxon:305"					
	/plasmid="megaplasmid"					
	/complement(147..1499)					
	/gene="RSp0310"					
	/note="RS05460"					
	/complement(147..1499)					
	/gene="RSp0310"					
	/function="cell processes; transport of small molecules"					
	/note="Product confidence : probable					
	Gene name confidence : hypothetical					
	predicted by Codon Usage					
	predicted by Homology					
	/codon_start=1					
	/evidence=not experimental					
	/transl_table=11					
	/product="PROBABLY MACROLIDE-EFFLUX TRANSMEMBRANE PROTEIN"					
	/protein_id="CAD17461.1"					
	/db_xref="GI:17430779"					
	/translation="MNAPRKQTRAEALRLRNASTTHRPANPCWIPDAAAGLHPGR					
gene						
CDS						

```

WRARFMAIFGQALSIGSALTOFVLLWMTDTTGSVSALATAGLAALLPOAVLSPLG
GTFADRVSRRLMIVAGISALCMILLIALFLTGRIALWHAYVMWAIWSAQAQAPAA
AAASVTLVPGFLVRAAGLNSQSLTLVVAAPLGAIGVMPGIALGGLDVATALL
GIAPLLCCRIPOACVSSQKTLWPAFREGVDLYMKTGFLRWLVLLGAVLITMPTF
TLLVLLVKTHGGAAGVAMLMGSLGIGMAGGLVVAAPRQVAVILCCFAMSCVVA
LALTALAPGSLFGVAVAWVTSIGTTFVFGNAPLTALLQTVIPNHLOQRVLSLNTWG
LAAPVGLALFTPLCEVIGVRAIPFAIGLGLGAASLAGFLSPSLRDLRRPSREARTD
"
complement (1531..2961)
/feature="oprM"
/feature="RSp0311; RS05459"
complement (1531..2961)
/feature="oprM"
/function="cell processes; transport of small molecules"
/feature="Product confidence : probable
Gene name confidence : putative
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE OUTER MEMBRANE CHANNEL LIPOPROTEIN"
/protein_id="CAD17462.1"
/db_xref="GI:17430780"
/translation="MRRLPLVLALSACSLPPALIKPALVPVATFAESGPFQOAH
ADLQWRTMGDRRLGLTGLANNRDLRLAALNVEVTRAQYSQRAARLPGLDAGTS
ASQRTQANDSPAALAAANSEQGVNVLNADIFLGRVRSLSDAAFARYLASDQ
RAAQISLVGAVADYFAERLAQOROLAHTLADWRQSLDLARRLKTAKQNSLDVA
QAEQVNAQADLEARTALAAQNSLQRLVAGLADLPAPLSLETQPVVTLQAPGL
PSELLRRPDIQAEHNLAAANADIGARAAPFRLSLTASLGASPMAGSLPGGHR
VWSPSQITQFQGRRLAEKRLAEKRSIAIEYRTIQTAFREVDAGLAGRATVS
RQIAQRAVASAERRVELSNLRAGVEGRLELLDSQRLYASRQALLDLROGAFSN
AVALKYKGLGLAETDVATDRVARK"
complement (2966..6115)
/feature="nxd"
/feature="RSp0312; RS05457"
complement (2966..6115)
/feature="nxd"
/function="cell processes; transport of small molecules"
/feature="Product confidence : probable
Gene name confidence : putative
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE MULTIDRUG EFFLUX SYSTEM TRANSMEMBRANE
PROTEIN"
/protein_id="CAD17463.1"
/db_xref="GI:17430781"
/translation="WPOFFIRPVPFAVIALFIILLGAIAIPKLPARYPSVAPPTVN
LSATYPQPTQMDGVLSLIERLSVKNLLYESSSDTSGSASIVATFKPGTDPM
AQVDQNKALPELPQVVRQNGLLTVSSAGSFLMIIGLQSDGDRDEFTALSDYMAR
NVBELRIEGRVQLQFCAQMRWDPKLIAYGLSMDLSAAISQQNAQIAPGS
VGASPALPQRTVPLTVQGLTTPQFPAALVLRANAGSKVMGDMVARVLQAQSYG
FINRENGKAATAAVQLAPGANAVRTAEQVQARMAELSRTPAGMSYSIPPTAPFVK
LSIEKVIOTLLEAVLFLVNLVFLQNVRYTLIPAIVAPIALITFTIMLAGSFSINV
LTMFGWLAIGIIVDDAIVVNVNVERLMAQGLSPRDATIKAMKEITGAVIGITVLT
AVFIPLMAGSGSVGIYQFTLSMAVSLVFSFLALTLTTPALCATLLKVPDPFHKHR
GVFGWNNRFRDRTQGYAARVGLVTRGRVMLVFGAIAAVLALFQLPSSFLPEED
QGYMTTQLPADATERTLDVKAERHVASRASIQSNMSIVGFGSGSGSNAAMAF
TMLKDWDRNAGTAEQNRMLAQAMAGATEGTVMSLPAPPAIDELTSGSFTLRQDRA
NOGDAALAAKTELLEAQSRIVSGVYDPGLPAGTSIRLDIQRKAEALGVATSTIS
BTLSTAIGSLVNDPBNAGRMQOILIQADAPRMQIDDLVLYRNAGGVMPLSEVV
RPVWSETPLQVRCGFPAPRIAGNAFPCASSGAGMAEMERLALQLPQGLAIATIGQS
LOERQSAQAQMLMALMVLVFLVLAALYESWAIPLSVMLVVPVGLGLAVLALLRGM
PNDFVFKVGMITVIGLSAKNAI LVEFAKQREQQGLVESAVTAAARLRLPIMLTSL
AFGLGVPLMTAMGASAEQTAIGTVGPGMVSAVTALAIFFVPVFFVFMVQLORFEA
WRASQAPPPVPHPQOEG"
complement (6136..7293)
/feature="nxd"

```

```

/feature="RSp0313; RS05456"
complement (6136..7293)
/feature="mexC"
/function="cell processes; protection responses;
drug/analog sensitivity"
/feature="Product confidence : probable
Gene name confidence : putative
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE MULTIDRUG EFFLUX SYSTEM PROTEIN"
/protein_id="CAD17464.1"
/db_xref="GI:17430782"
/translation="MKIERLRHGGVHVLLAAMLAGCGPSEPOSAPPAAEVATTVP
AHLELTEDLPGRVAARVRAEIRVSGIVCORRLPEQGTVEVRAGQPLQINAPPRADA
DTAARLRARAAALRARVQTTLRQLPLVEADAVSRQAYDDAVSQRDQAADVEQARAT
LARRQLDKFATVEAPIPRGIDQALVTEGALVSSGSDSNPMARIQQIDQVYDVRRPAS
SLAALRQTLATQKTGAGDGLPVAVLRNSGEPYAKGHILFSGINVDAGTGDVLLRVLV
NNPQRLPPGMFVARVPHARYTDALTVPQAVRVVGQPVQVMTLDANSHARKPVVEL
GELADRRYIRAGLSAQKIVVEGMRSLSDGAAVAADHMKSPPEVPAGPAR"
7392..8162
/feature="RSp0314"
/feature="RS05454"
7392..8162
/feature="RSp0314"
/function="miscellaneous; not classified regulator"
/feature="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE TWO COMPONENT RESPONSE REGULATOR
TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD17465.1"
/db_xref="GI:17430783"
/translation="MRILNORGAAVPMPPSPGSAQSAQALILIAEDEPEIADILAA
YLARSGLSRVHAADGREALHLHLKLDLVLDVQMPHVDGKALSIHRGDTFVIM
LTALDDQIDKLMGLRGADYVVKPPNPAEVAVARTQAVLRSMGSGTGRDORVLRVAP
FQIDLNHEATVQVQKHALALTLEFKLLAQARAKRVFVSRRAELLAACLPBGDTL
ERTVDSHISKRLKLEDLVGMVPGVGVGRVLSLE"
8159..9241
/feature="RSp0315"
/feature="RS05453"
8159..9241
/feature="RSp0315"
/feature="mexC"
/function="miscellaneous; not classified regulator"
/feature="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE TRANSMEMBRANE SENSOR KINASE
TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD17466.1"
/db_xref="GI:17430784"
/translation="MKLAGLSRQIALSMWAEQVTLVLTSAFYVYLAITYWSRDF
QASWLTGTEWVWLIATLGLAMAIIVANLARRILVNSVADSIIRVARDGLGV
RAIAGDSLGEAALLADDFNALQDQVRTEEQVFMAAIAHELRITPTVTLRGLQL
AEGVTPDEPQFRSLTQVESLTRLIEDLRVSLAESGHLDLQVRESDLVAEVKAVVA

```

Alignment Scores:  
Pred. No.: 4.13e-25 Length: 203050  
Score: 497.00 Matches: 114

Percent Similarity: 56.29% Conservative: 65  
 Best Local Similarity: 35.85% Mismatches: 13  
 Query Match: 24.28% Indels: 16  
 DB: 1 Gaps: 5

US-09-889-746-2 (1-400) x AB050700 (1-203050)

QY 29 GUGSerLeuProThrvAlaGluLeuGluProValAlaIleThrIleAspIleSerGlyMet 48  
 |||||  
 DB 122087 GAGGCGCCGCGCGCTGAGCTGAACTCGAGGCCACCGTGGTGTCCACCGACCGCAGC 122146  
 |||||

QY 49 AlaLeu-----AlaAsnArgIleThrGlnMetProHisThrIleValIleTyr 65  
 |||||  
 DB 122147 GACCTGCTCCGCGGACCGCGGCTCAGCATCAGCCAC----- 122188  
 |||||

QY 66 GUGGlnGlnIleGlnGlnAlaThrGlySerArgIleLeuAlaAspValMetAlaGln 85  
 |||||  
 DB 122189 ---GAGAGCTGGGCGACCTGCGGCGCTCGGCGACCTGCGGCGCGCTGAGCAGAA 122245  
 |||||

QY 86 LeuIleProSerLeuGlyValSerSerGlyThrThrSerAsnProGlyGlnThrMetHis 105  
 |||||  
 DB 122246 GTGCTGCGCGGCGGCTGCGGCGCTGCGGCGCTGCGGCGCTGCGGCGCTGCGGCG 122305  
 |||||

QY 106 GlyArgGlnValGlnPheLeuLeuAsnGlyValProLeuThrGlySerArgAspIleSer 125  
 |||||  
 DB 122306 GGGCGGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 122365  
 |||||

QY 126 ArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyAla 145  
 |||||  
 DB 122366 GCGACCTGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 122425  
 |||||

QY 146 ThrSerIleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrIleSerAsp 165  
 |||||  
 DB 122426 AGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 122485  
 |||||

QY 166 LeuGlnGlnGlnPheGlnThrArgIleGlyValHisGly-----SerIleLeuSer 183  
 |||||  
 DB 122486 GGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 122545  
 |||||

QY 184 SerGlnGlyIleGlyTyrGlnValGlyGlnSerValAlaGlyValSerGluAsnGlyAsn 203  
 |||||  
 DB 122546 GCGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 122599  
 |||||

QY 204 ValLeuAlaArgLeuAspValAspTyrArgThrThrGlyGlyAlaPheAspAlaAsnGly 223  
 |||||  
 DB 122600 GTGACTACGAATTCATATGTCGCGACGACGACATGTCGCGGCTCTATGACGCCCGCG 122659  
 |||||

QY 224 LysArgIleAlaProGluProAlaGlnThrAspIleGlnAspSerIleSerLeuSerVal 243  
 |||||  
 DB 122660 CACCGCATCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 122719  
 |||||

QY 244 AsnThrAsnValAspIleProGlnLeuAspAspIleAsnIleAsnLeuAlaLeuThrHis 263  
 |||||  
 DB 122720 GCGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 122779  
 |||||

QY 264 TyrAsnAspIleGlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeu 283  
 |||||  
 DB 122780 TAGGACCGGAGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 122833  
 |||||

QY 284 PheGlyGlnIleAspProSerLeuAsnAlaIleGlyGlyLeuSerLeuSerGlnIleProLys 303  
 |||||  
 DB 122834 CCGGCGGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 122893  
 |||||

QY 304 ThrThrIleSerThrPheAsnIleAsnIleAsnIleAsnIleAsnIleAsnIleAsnIle 323  
 |||||  
 DB 122894 ATCAAGAACCGGCTGCTCAACGTCGATCAGAACGAGATGCTGCGGCGGCGGCGGCGG 122953  
 |||||

QY 324 AsnThrAsnAlaTyrTyrArgArgGluLysGlyArgPheTyrProPheValAla 341  
 |||||  
 DB 122954 GGGGCGGAGTTCATACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 123007  
 |||||

RESULT 7  
 AB050700

LOCUS AB050700 2803 bp DNA linear BCT 14-JUL-2001  
 DEFINITION Vibrio parahaemolyticus *tutA* gene for ferric aerobactin receptor  
 precursor, complete cds.  
 ACCESSION AB050700  
 VERSION AB050700.2 GI:14717833  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 . Vibrio parahaemolyticus (strain:WPI) DNA.  
 Vibrio parahaemolyticus  
 Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
 1  
 REFERENCES  
 AUTHORS  
 TITLE  
 JOURNAL  
 Puhashashi, T. and Yamamoto, S.  
 Cloning and characterization of the *tutA* gene in *Vibrio*  
 parahaemolyticus  
 Unpublished  
 2 (bases 1 to 2803)  
 REFERENCES  
 AUTHORS  
 TITLE  
 JOURNAL  
 Puhashashi, T. and Yamamoto, S.  
 Direct Submission  
 Submitted (31-OCT-2000) Shigeo Yamamoto, Okayama University,  
 Faculty of Pharmaceutical Sciences; 1-1-1 Tsushima-naka, Okayama,  
 Okayama 700-8530, Japan  
 (E-mail: syamamoto@phs.u-oka.ac.jp,  
 Tel:81-86-251-8641 Fax:81-251-7926)  
 On Jul 13, 2001 this sequence version replaced gi:13928547.  
 FEATURES  
 source  
 1..2803  
 location/Qualifiers  
 /organism="Vibrio parahaemolyticus"  
 /strain="WPI"  
 /db\_xref="taxon:670"  
 180..185  
 misc\_feature  
 193..211  
 /note="putative Fur box"  
 202..207  
 -10\_signal  
 244..2421  
 gene  
 /gene="tutA"  
 244..2421  
 CDS  
 /gene="tutA"  
 /codon\_start=1  
 /transl\_table=1  
 /product="ferric aerobactin receptor precursor"  
 /protein\_id="BAB47148.1"  
 /db\_xref="GI:13928548"  
 /translation="MINKRFSILNLIIVLMSGGYAAHAAEAQQLASQNADEQLV  
 VGSWPKSIDIPGVFVQGEIRIAQVRAKRTLDIISATIPSIDVGTGNTGON  
 LRGRLMIVMGVLSQSRIPSLRDLADPNRIRIEVLSGATISYGAGTGVNII  
 TKKYSDFAPESFPGSGNSDDPQKAAQVAGNDIVKAGSVVSTGQAF  
 GNGDIVTDPDISQSLQYNSITDWSASAIQISDASKLVAVVQYSDSPGLIYV  
 SKFVVRKGFSDREHGTBRVLSANVYADNVFQHQLIGLSYRSDDFTFYVSS  
 QOETEVFAGRLALASKWGAFAVYGVADYLDLPDSNQLPFTTADNSGNLIRYAE  
 VGRYPGVNTSYAMFVQCDYOINODMSYQAGRYQYMDNKIDDFAVAVQKGLASGK  
 VSADVPFGSITDYVSLFNVTGIYQINDESGLMANFSQGFELADPAKYDGSGYEAVD  
 ANGHYALKDSINVDKSMGSKTDSYEIGYRLDPTVALQAAGYVSDNSIKYDKT  
 LLITTVDRQRYGAEANINWVSDDILKASGHVSELTNGKMEPLSAGKATSK  
 TNMAAYVDDTSVRLQSOTFEDVDDNNKLDGTTTDLGKRVNLPVGQLPFGIQL  
 IDEDTVMGGRAPQILYAHAYDAAYDYKGRRTFLNYSLEY"

BASE COUNT 766 a 589 c 686 g 762 t

ORIGIN

Alignment Scores:  
 Pred. No.: 2.65e-27 Length: 2803  
 Score: 495.00 Matches: 138  
 Percent Similarity: 49.76% Conservative: 70  
 Best Local Similarity: 33.01% Mismatches: 159  
 Query Match: 24.18% Indels: 51  
 DB: 1 Gaps: 10

US-09-889-746-2 (1-400) x AB050700 (1-2803)

QY 8 GlnTrpLeu-SerLeuProLeuLeuSerValAlaValThrGlnGlnLeuTyrArgAlaGln-- 26  
 |||||  
 DB 291 CAGTGATATGACGCGGACGACGATGGCTGACAGAACGCTCAGCAATGCGGCTCTCAAAA 350  
 |||||

QY 27 ---ProGlnGlnSerLeuProThrvAlaGluLeuGluProValAlaIleThrIleAspIle 45  
 |||||

```
Db 351 CCACAGATGAACAAATTAGTCGTGTT----- 378
Qy 45 sSerGlyMetAlaLeuAlaAsnArgileThrGlnMetProHisThrThrLysValIleTy 65
Db 379 ---GGCAGCGTCATGCCAAATTCATTCTGATATCCCGGTACGGTTTGGTTGTAGG 434
Qy 65 rCluGluGlnIleGlnGluAlaThrGlySerArgGlnLeuAlaAspValMetAlaGl 85
Db 435 GCAAGAAGAAATTTGCTCAACAATACCGACAGGTAAACTCTGGGTGATATTCTTTTCG 494
Qy 85 nLeuIleProSerLeuGlyValSerSerGlyThrThrSerAsnPheGlyGlnThrMetHi 105
Db 495 CACGATCCCTCTTTAGATGTTGGCAGCTGGCGCGGACAACTACGCCAAATCTCGC 554
Qy 105 sGlyArgGlnValGlnPheLeuAsnGlyValProLeuThrGlySerArgAspIleSe 125
Db 555 TGGCAGAGCCATGCTGCTCATGATTGATGGCTTCTCTGCAATCTCTGCTCTATCAG 614
Qy 125 rArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyAl 145
Db 615 TCGTCAATTTGGATGCGATTGACCCGTTCAACATCGAACGAATTTGAAGTCTTCAGGCG 674
Qy 145 aThrSerIleTyGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLysSerAs 165
Db 675 GACCTCCATTTATGGGCTGGTCCGAGGGTGTGTCATTACATCATCATCAAAAAGC 734
Qy 165 pLeuGluGluGlu---GlnPheGluThrArgIleGlyValHisGlySerLysLeuSerSe 184
Db 735 GTACTCAGATGAATGGCGCTTTGATCTTTTCTGCGCGGTACCTCAGGCTTCAACAGCAG 794
Qy 184 rCluGlyIleGlyTyGlyGlnValGlyGlnSerValAlaGlyValSerGluAsnGlyAsnVa 204
Db 795 CGATGATTTTGACTACAAAGTCGCAATCCGTTGCTGTT-----GGAAATTCATCGT 848
Qy 204 lLeuAlaArgLeuAspValAspTyArgThrThrGlyGlyAlaPheAspAlaAsnGlyLy 224
Db 849 AAAGCAGCTGGCTCAGTGGTGTTATCCGAGACACAGGTGCTATTGACGGAATGGTGA 908
Qy 224 sArgIleAlaProGluProAlaGlnThrAspLysGlnAspSerLysSerLysValAs 244
Db 909 CATTGTCAGCCGACAGATTTCTCAAGTTCTCTGCAATACAACTCAACGCTTGATGTTAT 968
Qy 244 nThrAsnValAspTrpGlnLeuAspAspLysGlnAsnIleAsnLeuAlaLeuThrHisTy 264
Db 969 GGGGTCCGAGAGATCCAAATCTCCGATCCGACGAAGTTAAACCTTGTGCGCAATACTA 1028
Qy 264 rAsnAspLysGlnAspThrAspTyAlaProAspTyGlyAsnArgLeuAlaValLeuPh 284
Db 1029 TGACAGCCAGCAGATTCGCCATC-----GGCTTGATCAT 1064
Qy 284 eGlyGluLysProSerLeuAsnAlaIleLysGlyLeuSerLeuSerGluGlnProLysTh 304
Db 1065 TGTTAATACCAATTTGTTGATGTGCGAAAGGTTCCGATTCGATCGTGAGCGGTAC 1124
Qy 304 rThrLysSerThrPheAsnIleAsnTyRhiHisAspAspLeuTrpGlyAsnThrIleAs 324
Db 1125 AGAGCGTGTACTTTAAGTGCACAACTACACACGACAACTGATTCGGTTCATCAATGAT 1184
Qy 324 nThrAsnAlaTyTyArgArgGluLysGlyArgPheTyProPheValAlaProPheSe 344
Db 1185 TGGTGATTTGTTTACCGGAGGAGATCAAACTTTTACACGCTATTACCAAGCTCTTC 1244
Qy 344 r-----IleAlaLysAlaLeuProIleLeuGlnInsMetAsnLeuPr 358
Db 1245 ACAACAAGAAACCGAAGTGTTCGCGGGCGTTTAGCGCTGGCGAAAGTTGGGGTGCATT 1304
Qy 358 oSerAlaThr-----LeuAspAlaTyR----- 365
Db 1305 CAGTCCCGTCTACGGTGTGGATGCATACCTAGATCGATTTCGATAGCAACCAAGCTTTGTT 1364
Qy 366 -----ThrLysAlaProGlnAlaArgAlaTyGlyValLeuGlnInsGluSerLysAl 383
Db 1365 CGACAAAACCATCGCAGATACTCA-----GGTACCTTAATTAACCGCAACGTATCG 1415
```

```
Qy 383 aGluValLeuGlyArgValProAsnLeuAsnLysProLysArgAlaLeuPhe 400
Db 1416 GGAAGTG---GGCGGTATCTCTGGTGTAGACGTAACCTCTTACGCGATGTTT 1464

RESULT 8
AB066099 5567 bp DNA linear BCT 12-JAN-2002
LOCUS AB066099
DEFINITION Vibrio parahaemolyticus genes for putative anaerobic
C4-dicarboxylate transporter DcuB, ferric aerobactin receptor
precursor IutA, hypothetical proteins, complete cds.
VERSION AB066099
KEYWORDS AB066099.1 GI:18157471
SOURCE Vibrio parahaemolyticus (strain:WP1) DNA.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE 1
AUTHORS Funahashi, T. and Yamamoto, S.
TITLE Cloning and characterization of the iutA gene in Vibrio
parahaemolyticus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 5567)
AUTHORS Funahashi, T. and Yamamoto, S.
TITLE Direct Submission
JOURNAL Submitted (12-JUL-2001) Shigeo Yamamoto, Okayama University,
Faculty of Pharmaceutical Sciences; 1-1-1 Taushima-naka, Okayama,
Okayama 700-8530, Japan
(S-mail: syamamoto@phesant.pharm.okayama-u.ac.jp,
Tel: 81-86-251-8473, Fax: 81-251-7926)
FEATURES
source
1. 5567
/organism="Vibrio parahaemolyticus"
/strain="WP1"
/db_xref="taxon:670"
complement(433..1764)
/gene="dcuB"
/gene="dcuB"
/gene="dcuB"
/codon_start=1
/transl_table=11
/product="putative anaerobic C4-dicarboxylate transporter
DcuB"
/protein_id="BAB83800.1"
/db_xref="GI:18157472"
/transl_table="MUYEFLLVLMYIGSYGGIGLVGVVSGILVIEFIFKMPPT
SPPTVMIIILAVVTCASILEAAGLKLQVAERVLKPKRVLTLPAPVTVYPMFL
LGTHAVYSIMPIIGDVALKNGIRPERPMAAASVASOIAITASPISAUVVYLAQLSD
IQHEITLLSILLVTVPATLFGTLLMSLYSIKRGKELEDDEYQERLKPVPVREKLANT
TATSDDEVLPFSARNVLLFIASILVIVIAMPPDIRTIYDGAKEPISNMAYIOMMLJC
FGGIIILATKTDPRDVENGVYFKSGMVAAIAIFGIAMSDTYFOYAMPQKSGIVEMV
TNPWTPALALFIVSVVNSQAATARMMLPVGLGLDLPALLIGLMPAVYGYFPINY
PSDIATVNPDTSGTTKIGKWFNHSVGLIGVIGACCLGYVLGQIIIPS"
2183..2917
/function="unknown"
/note="similar to alcaligin biosynthesis protein AlcD"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB83801.1"
/db_xref="GI:18157473"
/transl_table="MSESFPPQLPQTSQTTPYHGDIGRIESDAIHINDINPIVQE
LYRQLSETSEAGRAYWLTWDLCKQPIYVALVSYFGFSLPNLRNMAQYLOPCFV
RGFRPADQEHIQGSHITLIQAQGRQIRALDFYQEQNMWIRIRPQTHQLMSDGIILA
CLLRLOQHPQLTDAILKEHASLWLSALGLDERHIDALHGVPHEQPLKVRKSLVYK
CEGKLCEDCPRAAKNQLIMQKSN"
2957..2975
/misc_feature
/note="putative Fur box"
3008..5185
/gene="iutA"
/gene="iutA"
/gene="iutA"
/function="siderophore receptor"
```



```

/ codon_start=1
/ trans1_table=1
/ product="ferric aerobactin receptor precursor IutA"
/ protein_id="BAB81802.1"
/db_xref="GI:18157474"
/ translation="MINKRFPELNLIVLMSGGYAAHAAAEQAQQLASQNADEQLVIVV
VGSVWPKSISDIPTGVTVFVVGQBEILAQQRAGCTGLDILISATIPSLDVGTCGRTVYCGON
LGRRLMVLWIDGVSLQSSRRPILRQDLADIPNIEILVYLSGATSYGAAGAYINILIR
TKKASDELAFESFVGQSGSGFNSSDDPPYKPAQSAAGANDILVKRGSVYVSEEGADP
NGNDIVTPDILISQGLISQNSITLDVNSATDILSPASKNLVAQYVSDOODSPYGLIYINVA
SEPVNKRGSFDSREHGETRERVLISVYNDHNVFEGHQLIGELSTYSBDTPPYPIQSSSS
QOETVFERLRLAKSKWGFASVAVGVADVLDRFDSNQLFKRTLADNSGNILNFTVLE
VGRVYGVGVTVYAMFVQGVQDIQINDQSSVQAQRYQYDNKLDIPVLAASVQNIASGGG
VSADVVAGSGSTDSVSLFRVGVITVQINDQSSQIMAFSGOFEIADPARKYQSSGEAYAD
ANGHVALKDSINAVDSSKMSGILKTDISEYICGRYLDITPVVLAQAGVYSDNSIRKDKTR
LLITVWDEQRVYGAENAINSGIITDSEIDLLGASGHVYVLAATNGKSWMELISIXDSTK
TNMAAAVYQDDYSVRLSQSQETMPDEYDADNNTLLDGTTPDILIGNVNLPGVLGFEIQLNL
LDEDTYTAGGCAQQLIVLAHDAALAYDYKGGRTFTTLYNTLEY"
complement(5238..5420)

```

BASE COUNT	1555 a	1192 c	1334 g	1486 t
ORIGIN				

Alignment scores:	
Pred. No.:	6,28e-27
Score:	495.00
Percent Similarity:	49.76%
Best Local Similarity:	33.01%
Query Match:	24.18%
DB:	1
Length:	5567
Matches:	138
Conservative:	70
Mismatches:	159
Indels:	51
Gaps:	10

US-09-889-746-2 (1-400) X AB066099 (1-5567)

Oy	8	GlnTlPleuSerLeuProLeuleuSerValAlaValThrgIngluLeuYrAlaGln--	26
Db	3055	CAGTGGAAATGACGGCCAGCATTTGGCTGCAGAACCAAGTCAGCAACTGGCCTCCAAA	3114
Oy	27	---ProangIuSerLeuProThrValGluLeuGluProValValIleThrIleAspIly	45
Db	3115	CGCAGCAATGGAACATTATTCGTGGTT-----	3142
Oy	45	sSerGIyMetAlaLeuAlaAsnArgIleThrClmetProHisThrThryseValIlety	65
Db	3143	---GGCAGCGTCATGCCAANAATCCATVITCGATVITCCCCGGTAGCGTTGGTGTGTAAG	3198
Oy	65	rGlIuGluGlnIleGlnGlnAlaIleThrGlySerArgIngluLeuAlaAspValMetAlaG	85
Db	3199	GCAAGAAGAANAATGCTCAACAATAACCAGCAGAGTAAACTCTGGGTGATATCTTCTTCC	3258
Oy	85	nLeuIleProSerLeuGlyValSerSerGIyThrThrSerAsnPheGlyGlnIleMethI	105
Db	3259	CACGATCCCTCTTATGATGTTGGCATCGCCGCGCGCAAACTCAGCGCAAAATCTCGC	3318
Oy	105	eGlyVarGlnValaGlnPheLeuLeuAsnGlyValProLeuThrGlySerArgAspIleSe	125
Db	3319	TGGCAGAGCCATGCTGGTGTCATGATTAATGAGCCATTCTCTGCATATCTTCGTCCTACG	3378
Oy	125	rArgIInleuAsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGIyAl	145
Db	3379	TCGTCAATTGGATGGCATTTGACCCGTCACAATCGAAGCAATGAAGGCTTTCAGGGCC	3438
Oy	145	aThrSerIleTyrgIySerGIyAlaThrGlyGlyLeuIleAsnIleValThryseTrs	165
Db	3439	GACCTCCATTATTTGGGGCTGGCGACCGGTGTGTTCATTAACTATCATCATCAAAAAAGC	3498

QY 165 pleuGluGluGlu---GlnPheGluThrArgIleGlyValHisGlySerIysLeuSerSe 184  
Db 3499 GTACTCAGATCAAAATTTGGCGCTTGAATCTTTGTGCGCGGTAAGTCAAGCTTCAACAGAG 3558  
QY 184 rGluGlyIleGlyIyrGlnValIglYnSerValAglYalSerGluAengIAsnVa 204  
Db 3559 CGATGATTTTACACAAAGTCGACAAATCCGTTGGCTGT-----GGAAATGACATCGT 3612  
QY 204 lLeuAlaArgLeuAspValaAspIyrArgGthrThrGlyGlyValaPheAspAlaAengIly 224  
Db 3613 AAAAGCAGCTGCTCAGTGGGTATTCGAGACACAAGAGTGCATTTGACGGAATATGGTGA 3672  
QY 224 sArgIleAlaProGluProAlaGlnThrAspIyrGlnAspSerIysSerLeuSerValAs 244  
Db 3673 CATGTCAAGCCGACAAACATTTCTCAAGGTTCTGTCAATATCAACCAAGCTTGATGTAT 3732  
QY 244 nThrAsnValAspTrpGlnLeuAspAspIyrGlnAsnIleAsnLeuAlaLeuThrHisIe 264  
Db 3733 GGGGTCCGACAGATCCAAATCTCCGATGCCAGACAAAGTAAACCTTTGCGCGCAATACTA 3792  
QY 264 rAsnAspIyrGlnAspTrpAspIyrAlaProAspIyrGlyAsnArgLeuAlaValLeuPh 284  
Db 3793 TGACAGCAGACAAAGTATCCGATAC-----GGCTTGATCAT 3828  
QY 284 eGlyGlyIyrProSerLeuAsnAlaIleIyGlyLeuSerLeuSerGluGlnProIySth 304  
Db 3829 TGTATATAGCAAAATTTGTGTGTGCGCAAAAGGTTGCATTCGATCCGATCGTGCATGTAC 3888  
QY 304 rThlySerThrPheAsnIleAsnIyrHisAspAspLeuTrpGlyAsnThrIleAs 324  
Db 3889 AGAGCGTACTCTTTAAAGTCGAAACTACGACACGACGACAAAGTATTCGGTCAATATGAT 3948  
QY 324 nThrAsnAlaIyrIyrArgArgGlyIyGlyArgPheIyrProPheValaIaProPheSe 344  
Db 3949 TGTGGAATGTGCTTACCGCAGCAAGATCAAACTTTTACACGGATTTACCAAGCTCTTC 4008  
QY 344 r-----lIeAlaIyAlaLeuProIleuGlnSerMetAsnLeuPr 358  
Db 4009 ACAACAAGAAACCAAGTGTGTGGCGGCGCTTGTAGCGCTGCGAAAGTTGGGGTGCAATT 4068  
QY 358 oSerAlaThr-----LeuAspAlaIyr----- 365  
Db 4069 CAGTGCCTCTTACCGGTGTGAGTGCATACCTAGATCGATTGCATACCAAGCTTTGTT 4128  
QY 366 -----ThrIyAlaProGlnAlaArgAlaIyrGlyValLeuGlnSerGlnSerIyAl 383  
Db 4129 CGACAAACCAATCCGACGATACTCA-----GTTACTTAATTAACCAAGCTATGC 4179  
QY 383 agIValaIleuGlyArgValProAsnLeuAsnIyAspProIyAspArgAlaLeuPhe 400  
Db 4180 GGAAGTG---GGGCGTTATCCTGGTGTAGACGTAACTCTTAACGCGATGTTT 4228

RESULT 9  
AE013940  
LOCUS AE013940 13252 bp DNA linear BCT 26-JUN-2002  
DEFINITION *Yersinia pestis* KIM section 340 of 415 of the complete genome.  
ACCESSION AE013940 AEO009952  
VERSION AE013940.1 GI:21960333  
KEYWORDS  
SOURCE  
ORGANISM *Yersinia pestis* KIM.  
*Yersinia pestis* KIM.  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
*Yersinia*.

REFERENCE  
AUTHORS Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F.,  
Liles F., Petru N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,  
Strathearn J.D., Lindler L.E., Brubaker R.R., Plana G.V.,  
Foley S.C., McDonough K.A., Niles M.L., Watson J.S.,  
Battner F.R. and Perry R.D.  
Genome Sequence of *Yersinia pestis* KIM  
J Bacteriol. 184 (16), 4601-4611 (2002)  
12142430  
2 (bases 1 to 13252)

**AUTHORS** Deng, W., Burland, V., Plunkett, G. III, Boutin, A., Mayhew, G. F.,  
 Liss, P., Perna, N. T., Rose, D. J., Mau, B., Zhou, S., Schwartz, D. C.,  
 Fetherston, J. D., Lindler, L. E., Brubaker, R. R., Plana, G. V.,  
 Straley, S. C., McDonough, K. A., Nilles, M. L., Matson, J. S.,  
 Blattner, F. R., and Perry, R. D.

**TITLE** Direct Submission

**JOURNAL** Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445  
 Henry Mall, Madison, WI 53706, USA

**FEATURES** Location/Qualifiers

**source** 1..13252  
 /organism="Versinia pestis KIM"  
 /strain="KIM"  
 /db\_xref="taxon:187410"  
 67..2247  
 /gene="iutA"  
 /note="y3385"  
 67..2247  
 /gene="iutA"  
 /function="putative membrane; cell envelope: outer  
 membrane constituents"  
 /note="ferric aerobactin receptor precursor (cloacin  
 receptor); residues 9 to 726 are 66.25 pct  
 identical; residues 11 to 732 from GenPept :  
 >gb|AAD44750.1|AF141323\_21 (AF141323) iutA [Shigella  
 flexneri]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="aerobactin OM receptor"  
 /protein\_id="AAM86935.1"  
 /db\_xref="GI:21960334"  
 /translation="MKHRLVWLPCLVLMTPAAWEDQLVSNRSHRSVAENAOQT  
 TWVEGLEQVOQGLIKOILAQLIGDISQGRNYGNMHRGSIWMIDGVLRL  
 NSSRDSQLSDIDFNFAHLEVIIGATSLYGGSGTGLINIVTKKQGEQVLELQIG  
 KRLDVGHTNDENISAAAGSGTERAFGRFSYQRYGGWYDGGNEVLINDTOTGLQ  
 YSNRLDVGHTNDENISAAAGSGTERAFGRFSYQRYGGWYDGGNEVLINDTOTGLQ  
 LNSDRIPGTERHLINQVSNDFPQCQDLVAQVYRDESLTEYPPTLKDQKGVITIGAS  
 NVGRPYSTTNLAPFLDIDYINPFTLSGVRVYQYTENKVDVFGVAAQQAASGS  
 ATSDAPVQGVNNDNFNLAGLHLAHTSQOTWFFSQGFEPDLAKYGGSGYTLV  
 NHYQLQSNVNDKSLGIKVDYSYELGWRVYTGDLRTQMAGYLSLSDQTSINKTDM  
 TINVLPRKRIYGVGAVDYFPDENSEWAGATENLIKSETKVSGWKLTIDAAAPSK  
 ATAYIGAPGDWNLVQSQQNFDSKDKIDGNTIDELSSYALPVGKLSFSFENL  
 LDKYITVWQQAPLIY8FTYGSPLNLSYKGRGTGYNYSVLF"  
 complement(2506..3939)  
 /gene="y3386"  
 complement(2506..3939)  
 /gene="y3386"  
 /function="transport of small molecules; carbohydrates,  
 organic acids, alcohols, cations"  
 /note="residues 43 to 468 of 477 are 31.71 pct identical  
 to residues 29 to 459 of 463 from GenPept :  
 >emb|CAB12435.1| (299107) similar to H+-symporter  
 [Bacillus subtilis]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative carbohydrate symporter permease"  
 /protein\_id="AAM86936.1"  
 /db\_xref="GI:21960335"  
 /translation="WTGMSLDQSHLCSLKDEGDKLKIWMNNSGMKIPKELASVFGY  
 GMQCFSGLGVTFILFFYTDIMGISPVNA5MIFLIARWDAIHPLAGVMDTINMR  
 RKGPFYLLFTPLFLTVAAFYNEIASLMTKTIYAGVTIYILWGTLYALSDDIPWMS  
 STVMTEDEQERATKATCAMLGNAGIGATMILFFYISGLFAENSADRGYFAGVILMY  
 LGVIMLNGFMFKERVNVVTEKVTLCQTFIVWQNKPLFFILSAFFPMYFNSITVNT  
 FYIFFTNVMSDAELVSVGLITFTCALACIQTPELTHRFKRDLPFLCVLEIARV  
 GPMFTGYNNVMSVWLVITAIEMNTVPLISAMTADTVESYHYHTKRCIAITPSGO  
 TVGKLSVAVAGVSGLLSILGYMPNVAQ5MTWTLNGLFPFCISLLPAGVAVRILMR  
 KYKTEDEHALREELKQGRPHSSVGK"  
 4037..4717  
 /gene="y3387"  
 4037..4717  
 /gene="y3387"  
 /function="IS and transposon related functions"  
 /note="IS1661; residues 56 to 224 of 226 are 39.64 pct

**gene** identical to residues 1 to 167 of 173 from GenPept :  
 >emb|CAA63546.1| (X32970) orfA [Escherichia coli]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative transposase"  
 /protein\_id="AAM86937.1"  
 /db\_xref="GI:21960336"  
 /translation="MPATARYVTEVAMAEQAIPWYFGLTFLAVPPFORGNALNPRSW  
 IFLSMIRFGSMKHPFETRLAAVOHLYLSGKATLRETARQFSVGKSPLTRWIRAFRRQG  
 EAGLEHLSTRYTPFRLCVRYMNRCSAADA5AHFNIPNETIIQNMKRYREGGK  
 EALNPSKTGPTMPKDXVHDSKPFSEMTAELEKELEYLRAENAYLKKRKAALBEKAL  
 REQKKPE"  
 4147..5588  
 /note="insertion element"  
 /insertion\_seq="IS1661"  
 4771..5556  
 /gene="y3388"  
 4771..5556  
 /gene="y3388"  
 /function="IS and transposon related functions"  
 /note="IS1661; residues 1 to 257 of 261 are 49.22 pct  
 identical to residues 21 to 276 of 281 from GenPept :  
 >gb|AAK29066.1|AF345899\_2 (AF345899) OrfB [Klebsiella  
 pneumoniae]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="putative transposase"  
 /protein\_id="AAM86938.1"  
 /db\_xref="GI:21960337"  
 /translation="MARSTVYHASKPDGVIDDYADAVAKAIGALSRRHAORYGYRRWT  
 VALRKEFTLNHKTRKLMQHGLLSIRKRYRYRADGGRASDNLNLRNFTSEISG  
 LKMDVTFRVGAQKLYLQDLFNNETISWHMSRAALILCTKLEKAKVKYGRK  
 EGMLEHSDQGHVYTPMWSLMEAGTIRSMRSGNCLDNAMVENFFSHLKAEMYHRK  
 KYDSATLKRDIVYIHYINTERISLKTGMSPAEYRTQVEKQ"  
 complement(5740..8106)  
 /gene="y3389"  
 complement(5740..8106)  
 /gene="y3389"  
 /function="unknown"  
 /note="has GDEF sensory domain; residues 14 to 778 of 788  
 are 43.72 pct identical to residues 75 to 838 of 864 from  
 GenPept : >gb|AAG05460.1|AE004634\_3 (AE004634) conserved  
 hypothetical protein [Pseudomonas aeruginosa]"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical"  
 /protein\_id="AAM86939.1"  
 /db\_xref="GI:21960338"  
 /translation="MQQDSLKTRIKONAFWGDAYQHLHVMDTWAFISQNMGFSLYK  
 DFSLEGVFIIDGQKTRYVINGQLVNTDLSWLKDOIPLIDIRARAEKPYIAVDT  
 LEIAQGPALIAAALTPEGDPVQWVAGPSPVLFINILTPAELSAIGENYGINLRM  
 PRNSQDATSKPSLVIASDSTLTLDHWDQMPGTPLNILLPLALVALIIGLTAWLV  
 IRRAI5EAQATDRSRAALAABEERFHVAAESADMLWETAELQVTVLSPRFLIISGW  
 SVDHMLGNDLSLNCDTISLRWLQHTLTDNRNLOCTYFSAGSKRVCRIHAKPI  
 IQQGEIVGFRGTASDLTREIAEKRIEHLSDHLDLGTPLNLTMTFLENKLNLAAY  
 EHPVLINVLNLDKFPVNDIFGHVTDQVLYQVAERLSRCLDRQSVRQGGDFVLI  
 ISNLSSEYIELLCARVITRQTPVINDQBIYLGASIGITLAPQDSMQAELLERFAD  
 IAMEAKNRPNRMRVASEMDRLMORIELEKFLRLAVKSEFELYQYPRYTDMMQ  
 LGAEALVRNHPVLGLLMDPQFALAEETGLTATISDWTLLQACQADMTWPTSLIVS  
 VNISAIFRGCCLIERVQALLVTGLPSHRLLEITERIMIEDADGALKINTALKTLG  
 ILLIDDFGQYSLSNTHRRFPDGLKIDKFSIDKLTESHGQSVI6EIGINHGAISM  
 TVIAEGVETAPQLYLSQSLHCDVQGYFLAKPMPEALSLSVFSDDLVS"  
 complement(8110..8226)  
 /gene="y3390"  
 complement(8110..8226)  
 /gene="y3390"  
 /function="unknown"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hypothetical"  
 /protein\_id="AAM86940.1"  
 /db\_xref="GI:21960339"  
 /translation="MALLVGMTIIIAVGIIYIASQLNEQASVQSRFLIEKA"

```

gene      8198..8371
          /gene="y3391"
          /locus_tag="y3391"
          /function="unknown"
          /codon_start=1
          /transl_table=11
          /product="hypothetical"
          /protein_id="AA086941.1"
          /db_xref="GI:21960340"
          /translation="MIVPTSKAMVCIRPAKTRSPVIRVDGGLAVIARLSYD
          IGLKFNKRVK"
          /gene="y3392"
          /locus_tag="y3392"
          /function="putative transport; protection responses:
          drug/analog sensitivity"
          /note="possible multidrug efflux system, periplasmic
          component; residues 49 to 421 of 425 are 37.53 pct
          identical to residues 18 to 381 of 388 from GenPept :
          >glAK65569.1 (AE007277) putative drug resistance
          protein [Sinorhizobium meliloti]"
          /codon_start=1
          /transl_table=11
          /product="putative drug efflux protein"
          /protein_id="AA086942.1"
          /db_xref="GI:21960341"
          /translation="MFTLMWVSTPIFYVPRTRLSRLIRSPRMGRPAKORHRLS
          HALFLIGVTSATAPLPAVTAVAVSTIPDAVQIGIEAIQAVDTTRREG
          IARRLETBKMKGSELVEIDPALHQAQVQAQQLDSATSAHQAQVNLQRLG
          NNRYSQAEVDAQAQRDISRAVAQAQNLQIQQLSFTQIHAPISQMGSHFNV
          GSLINPAGSTLVNIYQIDPIRYVALINERDITATYQRAVAREKNIENLLIGTFEP
          RQLNGKQYPSGVPSVDNIDIKOTGVIIIRTPANRHLLIGVAVNVSLADDA
          KRYTVIRALQONKQGHVLYVTDKAVERAPVILGSEFEOQYAVTDGVKGERVIV

```

## Alignment Scores:

```

Pred. No.:      3,69e-26      Length:      13252
Score:          491.00      Matches:      123
Percent Similarity: 47.88%      Conservative: 69
Best Local Similarity: 30.67%      Mismatches: 155
Query Match:      23.99%      Indels:      54
DB:              1          Gaps:      9

```

US-09-889-746-2 (1-400) x AB013940 (1-13252)

```

QY      1 MetARGHISerHisTyrPheGlnTTPLeuSerLeuProLeuLeuSerValAlaValThr 20
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      67 ATGAAACACAAACACCTC-----TGGGTATTAAATCCGTGCTGCTCATGCTCACA 120
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      21 GlnGlnLeuTyrAlaGlnProAenGlnSerLeuProThrValGlnLeuGlnProValAla 40
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      121 CCAGCGCGCGTGGGCAAG-----GATCAACTTGTG 150
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      41 IleThrIleAspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThr 60
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      151 GTTTCAGCCCAACCGCTCACACCGAGCGTGC-----GAAATGGGCAACA 198
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      61 ThrLysValIleTyrGlnGlnIleGlnGlnGlnAlaThrGlySerArgGlnLeuAla 80
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      199 ACTTGGGTGTTTAAAGCGCCGGAACCTTGAACAAGTTCAAGCGCGCTCTGTAATCA 258
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      81 AspValMetAlaGlnLeuIleProSerLeuGlyValSerSerGlyThrThrSerAsnPhe 100
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      259 GACATCTTGCTCAATGATCCAGGATGATGATGATGATGATGATGATGATGATGATGAT 318
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      101 GlyGlnThrMetHisGlyArgGlnValGlnPheLeuLeuAsnGlyValProLeuThrGly 120
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      319 GGCATGAACATGCGTGGCCGCTCAATCATGATGATGATGATGATGATGATGATGATGAT 378
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      121 SerArgAspIleSerArgGlnLeuAsnSerIleAsnProAenGlnValAlaArgIleGln 140
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      379 TCACGTAGCAGACGCCCAATCGATTCATGATCGATTCGTTAATATTCACACATTGAA 438
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

```

QY      141 ValLeuSerGlyAlaThrSerIleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIle 160
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      439 GTCATCTCCGCGCGACATGCTTATGCGCGGGGTAGACCGGGGGCTTATCATATAC 498
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      161 ValThrLysSerAspLeuGlnGlnGlnPheGlnThrArgIleGlyValHisGlySer 180
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      499 GTCACCAAAAAGAGCCAGAGAAAGAAACAGTTGATGATGATGATGATGATGATGATGAT 558
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      181 LysLeuSerSerGlnGlyIleGlyTyrGlnValGlyGlnSerValAlaGlyValSerGly 200
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      559 TTTAACAGCCATTAACAGCACCATGATGATGATGATGATGATGATGATGATGATGATGAT 618
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      201 AenGlyAsnValLeuAlaArgLeuAspValAspTyrArgThrThrGlyValAlaPheAsp 220
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      619 -----CGCGCATTCGCCCGCATTCCTCTCTGTTCTTATCAACGTTATGGGGGTGTAAGC 672
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      221 AlaAsnGlyLysArgIleAlaProGlnProAlaGlnThrAspLysGlnAspSerLysSer 240
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      673 GGCAAAAGCAATGAGGTTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 732
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      241 LeuSerValAsnThrAsnValAspTyrGlnLeuAspAspLysGlnAsnIleAsnLeuAla 260
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      733 CTGACAGTATGAGGACCGGAACTTCATATATGACGAAACACCAATATGACGCTAAC 792
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      261 LeuThrIleTyrAsnAspLysGlnAspTyrThrAspTyrAlaProAspTyrGlyAsnArgLeu 280
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      793 ACTCAGATTTTCAACAGCAATCCGATGCAACACCGCTTATCTGGGCAAAATCTC 852
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      281 AlaValLeuPheGlyGlyLysProSerLeuAsnAlaIleLysGlyLeuSerLeuSerGln 300
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      853 TCAGAGTAAACGGGTACCGGACGAGCTTCACACG-----GCAGCGCTTAATTCAGACCG 909
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      301 GlnProLysThrThrLysSerThrPheAsnIleAsnTyrHisHisAspAspLeuTyrGly 320
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      910 ATTCGGGTACCGAACGCCATCTCATCACTGCAATATCCAAACCGGATTTCTGGGGG 969
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      321 AsnThrIleAsnThrAsnAlaTyrTyrArgAspGlyLysGlyArgPheTyrProPheVal 340
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      970 CAGATTTAGTGGCGCAAGTATTTACCTGATGATGATGATGATGATGATGATGATGATGAT 1026
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      341 AlaProPheSerIleAlaLysAlaLeuProIleLeuGln----- 353
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      1027 -----CCACACCTGAAGATGCAAGGTGACACTATC 1059
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      354 -----SerMetAsnLeuProSerAlaThr 361
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      1060 GGTGATTCACAGCAAAACCGATTTCTACGGCAGTAACTGACATTTGAACGTGAACCT 1119
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      362 LeuAspAlaTyrThrLysAlaProGlnAlaArgAlaTyrGlyVal-----LeuGlnSerGln 380
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      1120 ATGATAGCTTACA-----CTGACTTACGATGATGATGATGATGATGATGATGAT 1164
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      381 Ser 381
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      1165 AGT 1167
          |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

RESULT 10  
 AJ414145  
 LOCUS  
 DEFINITION Yersinia pestis strain CO92 complete genome; segment 5/20.  
 ACCESSION AJ414145 AL590842  
 VERSION AJ414145.1 GI:15978904  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Yersinia pestis.  
 Yersinia pestis  
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 Yersinia.

REFERENCE  
 1 (bases 1 to 208050)  
 Parkhill, J., Wren, B. W., Thomson, N. R., Titchell, R. W., Holden, M. T. G.,  
 Prentice, M. B., Sebatista, M., James, K. D., Churcher, C., Mungall, K. L.,  
 Baker, S., Basham, D., Bentley, S. D., Brooks, K., Cerdano-Tarraga, A. M.,  
 Chillingworth, T., Cronin, A., Davies, R. M., Davis, P., Dougan, G.,

Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Leather, S., Kariyeh, A. V., Moule, S., Oyston, P. C. F., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S., and Barrell, B. G. Genome sequence of *Yersinia pestis*, the causative agent of plague Nature 413 (6855), 523-527 (2001)

21470413  
MEDLINE  
PUBMED  
11586360  
2 (bases 1 to 208050)  
Parkhill, J.  
Direct Submission  
Submitted (04-OCT-2001) Submitted on behalf of the Versinia sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk

Notes:  
Details of *Y. pestis* sequencing at the Sanger Centre are available on the World Wide Web.  
(URL, [http://www.sanger.ac.uk/Projects/Y\\_pestis/](http://www.sanger.ac.uk/Projects/Y_pestis/)).

## FEATURES

source  
1..208050  
/organism="Yersinia pestis"  
/strain="CO92"  
/db\_xref="taxon:632"  
/note="Biovar: Orientalis"  
375..1151  
/gene="YPO0831"  
375..1151  
/note="YPO0831"  
/note="Similar to *Escherichia coli* galactitol utilization operon repressor Gatr SW:GATR\_ECOLI (P36930) (259 aa) fasta scores: E(): 2.9e-27, 39.0% id in 254 aa, and to *Escherichia coli* putative aga operon transcriptional repressor Agar SW:AGAR\_ECOLI (P42902) (289 aa) fasta scores: E(): 0, 46.7% id in 255 aa"  
/codon\_start=1  
/transl\_table=1  
/product="DeoR-family regulatory protein"  
/protein\_id="CAC89678.1"  
/db\_xref="GI:15978908"  
/db\_xref="SPTREMBL:Q8ZH09"  
/translation="MKAAIERRMIEFNIVTQOGKARVEDLAERFNVSSVTIRSDLSFL EKNQIVRSHGSAIPNTGVIAELTVHEKRRNSGVKSLIGQAAKLINSQGVTLVDSG TTTREIALRSVENVMTNGLDVAMELASAPGVETVLISGVLRKKSAMSPSGSOAE SLKRFPRKVLGVDGLRVGITYNEQESLRLMCDISEQIIAVADSTFKGRSC HMIREFGIDILVTDSDIPSEYVORLDMRMEVIAAEK"  
393..1073  
/gene="YPO0831"  
/note="Pfam match to entry PF00455 deoR, Bacterial regulatory proteins, deoR family, score 254.40, E-value 1.5e-72"  
435..500  
/gene="YPO0831"  
/note="Predicted helix-turn-helix motif with score 1114 (+2.98 SD) at aa 21-42, sequence ARVEDLAERFNVSSVTIRSDLS"  
1201..2499  
/gene="YPO0832"  
/note="agaZ"  
1201..2499  
/gene="YPO0832"  
/EC\_number="2.7.1.144"  
/note="Similar to *Escherichia coli* putative tagatose 6-phosphate kinase Agaz SW:AGAZ\_ECOLI (P42903) (425 aa) fasta scores: E(): 0, 63.6% id in 423 aa, and to *Streptomyces coelicolor* tagatose 6-phosphate kinase Agaz TR:O50522 (EMBL:AL009204) (435 aa) fasta scores: E(): 0, 57.2% id in 418 aa"  
/codon\_start=1  
/transl\_table=1  
/product="putative tagatose 6-phosphate kinase"  
/protein\_id="CAC89679.1"  
/db\_xref="GI:15978906"  
/db\_xref="SPTREMBL:Q8ZH08"  
/translation="WHPLEKLVORHKYAGQIGIVSVCSAHPLVTEAALEQALENDSSEL LIATNSQVDQFGGVTGTPVDFRDPYPAKADALHPLEKILIGDLGHPNRWQLPA

NEAMAHADDLIAAYVAAQPKXIHLDSCMSCADDPVPLSDSEIVAEARAARLAIAEATY CHENTHATVAVIGTEVPIGGATEEDELDEIVTSPAAKKTIECHQKAFESGVAECPW KGFALVQVGFEDHTGVIDFCPEKAHLSKVNDEYDQVFEAHSITDQTDAYRALV KHFPAIKVGFALTALREALFIEELFPAEQCSRLRQVQMEDVLMKSPYWKRY YLGDTEOVLRLARVYSFSDRMRYWPMEEAVQAAQKLFANLSAPIPLPLLSQYMPRQF TAVREGTSLFEATALVKQIKNTLIGYVNACHLFGNTSK"  
2496..3662  
/gene="YPO0833"  
2496..3662  
/gene="YPO0833"  
/note="Similar to *Drosophila melanogaster* glucosamine--fructose-6-phosphate aminotransferase Gfat TR:Q9V093 (EMBL:Y18628) (694 aa) fasta scores: E(): 0.00036, 22.3% id in 363 aa, and to *Escherichia coli* putative tagatose-6-phosphate ketose/aldehyde isomerase Agas SW:AGAS\_ECOLI (P42907) (384 aa) fasta scores: E(): 0, 48.0% id in 373 aa"  
/codon\_start=1  
/transl\_table=11  
/product="putative phosphosugar isomerase"  
/protein\_id="CAC89680.1"  
/db\_xref="GI:15978907"  
/db\_xref="SPTREMBL:Q8ZH07"  
/translation="MTQYLGYSREELTRLHASHTAOEIQOCPACWGTORIVRESOAA INDFLQALLAASNTRIIFTGAGTSAFAGRALAPLLAPLSHRLEAINTDLVATPEQY FSDIPITLLVSPARSGNSPESVAALQVASACLRCFHLVLTQGGOLYQYQATPAGA LAVLPAETNDSQALMTSSFSMMMAAMAFAGIDKYKSEIERFCGSADMFKEINHK IIDNTKSSQYQVRVYILGSGQLGAEALKELTAGKVAVATDTPLGRHGPKSIV NPDTLVVVISNDPYTRYOYDULLAEVRANNAVGVIAIAAQSDAVYTAGDHLVPGM AESSDGLFAYLMPFAQSPAFHSSVALNNTPDSPSPTEINRVQGVTIYPPKK"  
2646..3083  
/gene="YPO0833"  
/note="Pfam match to entry PF01380 SIS, SIS domain, score 42.30, E-value 1.1e-08"  
3159..3587  
/gene="YPO0833"  
/note="Pfam match to entry PF01380 SIS, SIS domain, score 45.70, E-value 1e-09"  
3680..4177  
/gene="YPO0834"  
/note="manX"  
3680..4177  
/gene="YPO0834"  
/note="Similar to *Vibrio furnissii* phenolpyruvate-dependent sugar phosphotransferase system (PTS) permease for mannose subunit IILMan C-terminal domain, ManX TR:P96162 (EMBL:U65015) (157 aa) fasta scores: E(): 6.8e-31, 51.6% id in 159 aa, and to *Escherichia coli* PTS system, N-acetylglactosamine-specific IIB component 2 Agav SW:PTPV\_ECOLI (P42904) (157 aa) fasta scores: E(): 8.1e-29, 48.1% id in 158 aa"  
/codon\_start=1  
/transl\_table=11  
/product="putative PTS transport protein"  
/protein\_id="CAC89681.1"  
/db\_xref="GI:15978908"  
/db\_xref="SPTREMBL:Q8ZH06"  
/translation="MTRPNVTRIDERLLHGOIRITWGKHTANLILVANDDAADGPN NSAFMOAGMKASAGGEVAFFSIQKTIDVIHKASPOKVPILCNSPTDVARLVGGV PITQCNGNNHYHDKRQIAKTYSVDEKLDJDAQFMAALGACTVQNTPTQTPINVIE LAARA"  
4224..5012  
/gene="YPO0835"  
4224..5012  
/gene="YPO0835"  
/note="Similar to *Vibrio furnissii* phenolpyruvate-dependent sugar phosphotransferase system (PTS) permease for mannose subunit IIPMan, ManY TR:P96163 (EMBL:U65015) (258 aa) fasta scores: E(): 0, 41.2% id in 262 aa, and to *Escherichia coli* PTS dependent N-acetyl-galactosamine-IIC component Agav TR:Q9K103 (EMBL:AF286498) (259 aa) fasta scores: E(): 0, 41.2% id in



```

Escherichia.
1 (bases 1 to 3668)
SmaJb,D., Smarda,J. and Weinstock,G.M.
Direct Submission
Submitted (16-JUL-2001) Human Genome Sequencing Center, Baylor
College of Medicine, One Baylor Plaza, Alkek N1519, Houston, TX
77030, USA
FEATURES
source
    location/Qualifiers
    1. 3668
    /organism="Escherichia fergusonii"
    /strain="BF873"
    /db_xref="taxon:564"
    466..1611
    /gene="flhB"
    466..1611
    /gene="flhB"
    /note="flagellar protein; similar to Escherichia coli
    FlhB"
    /codon_start=1
    /transl_table=11
    /product="FlhB"
    /protein_id="AAL01537.1"
    /db_xref="GI:22034295"
    /translation="MSBENDSEKTEDTPHRKQKAREGOVPRSRRLTSLSMLLAGHG
    LLLVGNVLAQLOHLNGHGLTFFRLVSDQPHMLHAAALLGAFILAILPFGGLFI
    TALVTPSIIIGGLNSAKAIKFNKLKJNPLSGIKRIFSGQMVSEMKVILKVLGAGIG
    GLFLWLNKAKFINLIFEPGLMALADISALLIGMLVILSLIPMAFDVIYQLWSNFK
    KLRMSLKEIKDEFNQEGDPOIKARVQLORQMAQORMTDVPADLVVNNPHTYSVA
    LKYOEGAMGPMVAVSGGLIALRIELAEENRVPMLEAPLARALYRHCDCPGTPVPA
    ELYSAVAELAVVYGLKWRKGYGSRPLTPKDLVPATMDFVQESRE"
    1611..3704
    /gene="flhA"
    1611..3704
    /gene="flhA"
    /note="flagellar protein; similar to Escherichia coli
    FlhA"
    /codon_start=1
    /transl_table=11
    /product="FlhA"
    /protein_id="AAL01538.1"
    /db_xref="GI:22034296"
    /translation="MANIATKLRLPNMKETQVQLAGPVLIMTILAMMILPPTFVLVD
    LFTFNIVLAILMILVAMFTQNTLEFSAPTVLFLSTLLRLANVASTRYLIMNGHTG
    SEAAQGVVEAFHFLVGNFAIGIIVFVILIINFVITKAGRIAEVGFARFLDGMF
    GKMAIDADNAGLIGEDAEKRRKEVTQEADFYGSMDSKFKVRGDAIAGLLIMAIN
    IIGLVIGVMOHDMTVGLAGETTYTLTIGDGLVAQIPGLVISTAAGVVVTRVANDQV
    GEQVVGQLTSPRIVLAAGVIGLMGMPGNFVELLFTVTLAIAWLRGRGETN
    KSTAAQNSDITDAEVAQNEATSDVQMEDMLGLEVGYRLIPMDHEQQQLTRVR
    GIRKFAQMGFLPPAVHIRNLDLAPHYRLIKLVGNIGHGEVQPRKMAINFQCAE
    GEVFTGPTCTPGLPALWIDVHRELAQTLYTVDPSSVVATHLNLHLSITHTDLEF
    GRQAQQLDQIDARQSPKLIIDDLIPAVISLTLTKVQLNLSRVPRIIDMTIITDLA
    EFATAQDADSLTARVRLSRATIQWFPGEDDIQLGLDLSLEQLLIQATQSGSAI
    EPGIAENLMKOTKALLHQEGIGAPPVLLVNPALRLMLSRVRLRIFPQLAVLSQSEIN
    TQNRVRYTLIGGKG"
    3701..4111
    /gene="flhB"
    3701..4111
    /gene="flhB"
    /note="flagellar protein; similar to Escherichia coli
    FlhB"
    /codon_start=1
    /transl_table=11
    /product="FlhB"
    /protein_id="AAL01539.1"
    /db_xref="GI:22034297"
    /translation="MSKPYLAGLLISPWALASGSWSSESFGCTWTRGQGSVKSKPL
    QSPGLPGARATTVHWIKITDGLTPSLGRIRLCNAKCVRLAGWAGEMPLPAGIAPE
    GPFRFEHASEAGTLRSPTVILSNQLTVSYHSSH"
    complement(4151..6289)
    /note="orf4; similar to Escherichia coli YejO"
    /codon_start=1
    /transl_table=11
    /product="unknown"
    /protein_id="AAL01540.1"
    /db_xref="GI:22034298"
    /translation="MKPLVQYSLRRAILVAMLVGTGTASATTLIDKDTFIDANSPVN
    ENRIENGATLSVNGATTGNTISDGLHANNGSVNVLYGEYGATADIADATLKTAAL
    QTSRSRTVIEQLSIFWNLISANSISAOQFYAAGAOQSTVEKGYPEOTERMVAGAVY
    LTSSEGIPTDSVTKIGIKTAVYLEGNSVDLIRSSIWATEGSAFDIARGNQKILQGS
    ASSQTGELLVARTGSSVNLALDTSHAAGDIRAEGSAIVDVALTNGSVLSGGMVNSL
    MLGNASLYEMTSSDITLSMAGMVKFAPAEAGEHHTLTLSLEGNGHFMNIDTEN
    GSSDLIDVTGHAEGSHTLAASGREANTDPLTLVTTGGGDAEFALNGRVDWMAWQ
    QELVRNQNQWELVQAKGATASDTAMLASAPQFMHQGLKVLRSRLAVKDTDSAAQ
    VMGAMLYNRSDIEGAQGSAYRLTONGMILGSKDVTGTACELTTGAYLSOSTGNIKHA
    RGGDSKVTGYGGGVYARLSAENGAVASTSAOIGHFSNKLNRMSDGGGTARADNBSLGY
    GLMVEYGRQITVGETTSLTPFVGUNAWLTPSENVTLNSGMTAETGDNRSLLAAGVRM
    STRLDVAEMAVTPYVTLVQOQLVKSSTOINDAYDTNDTGTGTGGTGIAGIYAIAP
    TASVWLEAEYKGESETESALAGNVGRVAP"
    complement(6370..6822)
    /note="orf5; similar to prophage CP-933X hypothetical
    protein"
    /codon_start=1
    /transl_table=11
    /product="unknown"
    /protein_id="AAL01541.1"
    /db_xref="GI:22034299"
    /translation="MDLNRASLDAAARVCRDRIKESILHDVENPPPSHLHQDFDSVS
    LFILHCLPCTMKREKSPFIRNAATAALNDEGLFGATILGDDIGHNAPGRKLMAYNNK
    GIFSNRTDIDDIQTILSEAPHQVDIKVEGKVALFKAANKKTGRIPVS"
    complement(7138..8343)
    /gene="traF"
    complement(7138..8343)
    /note="traF"
    /codon_start=1
    /transl_table=11
    /product="traF"
    /protein_id="AAL01542.1"
    /db_xref="GI:22034300"
    /translation="MNNKLLTRSLIAAGLIAAIPIYASASSNGFDARSSAMGGVGVASA
    HFGAPLTNPALLATREQDKISLIAPSIGAQSADPANLIDGFDVTTAWDNLENALG
    SGNEAAGKADSVAGAGEHANVGLSMVLAVPDELFPVALSVNSWAKGHARALV
    SQSDLDYLDVAVKGIIPGKDLDKLTSAEGMAALVTEYGTVAHPFTLQELPVGVG
    ITPKIORIETWYNVAINVYDSSDLRGNWQHOTSANIDAGFFASVTPSWMVALSAQ
    NLENKVKTYINGYQPAFIIRBELTAGTAMNERNVTLSADIDLTPSVNQYVDKNOY
    AATGAERUADWVQLRAGRYLDMRGNDRRVTVTGVGLSAGEAQFDLTAWAGDRRTIG
    GVAQPTFFH"
    complement(8694..9065)
    /gene="gp37"
    complement(8694..9065)
    /note="gp37"
    /codon_start=1
    /transl_table=11
    /product="gp37"
    /protein_id="AAL01543.1"
    /db_xref="GI:22034301"
    /translation="MLFRKIAFVAVPLFSSVAVADVAGAPCCPFDENLKENTKPLDH
    SLEKILKGVSYTWKEDKTGVLIAQDVKEVYVELVTKGRIKQVDYQKLVAPLIE
    AVREQQNEINALESRRRPMCT"
    9642..10439
    /note="orf8; similar to Escherichia coli hypothetical
    protein B1976"
    /codon_start=1
    /transl_table=11
    /product="unknown"
    /protein_id="AAL01544.1"
    /db_xref="GI:22034302"
    /translation="MIKMPKMTNDAGNMALPWDEALAIPLVLANLQDESKLVOLAD
    RFQQKRLVFLQGFDELPLKRIALLFCPLVLELQDEWLDGHEVLIYPAFPVDBE
    WQDDFGVLHNRVQVSSQWQGPITLNLWDIQDSFDSAGFNLIITHEVAHKLDFRNGD
    RAGVPLIPLREAVAGWEHDLHAADNDIQDEIDLVGESAASIDAYAAATPAECPAVLSE
    YFFSAPELFAPRFPALWQRCQFYQQDPLLRLAQNEPDAGNSSHQLH"
    complement(10436..11647)
    /gene="shp"

```

```

CDS
complement(10436..11647)
/gene="shf"
/notes="similar to Shigella flexneri shf protein"
/codon_start=1
/cranet_table=11
/product="Shf"
/protein_id="AA01545.1"
/db_xref="GI:22034303"
/cranelation="NRTYPTLTNTLPKRNPLACAGLIGIGNGILVLMPLQVTL
TGLSLVAGLGMFGSMPLPASPMWGRSEHGCKVWQASLGGTAFVMAVVM
AMAGMSAVNKLGLILSRITGLVSGVLAQVMAIQAGLDRMAALATISSGL
SGRLGPPPLAAMLSVNPVAPLMLAIFLALDLVLEVPPLPIAHQATVQV
SNLPLVALLLALVSLMOIGLSPHLSLDGNPREVSHVALLSLAALTAQF
LVRRPQHSPTVTLCTAAVAMTGLKICLIGLVVGVGIVMTSLGAAMATGYDILL
NRLITGKAGAVIATSHITGVGSLVLPVTRLPGEVFLTQAAVMALPLVMSLW
BALNRTHVDS"
11786..13528
/gene="lucA"
11786..13528
/gene="lucA"
/notes="Involved in aerobactin synthesis"
/codon_start=1
/cranet_table=11
/product="lucA"
/protein_id="AA01546.1"
/db_xref="GI:22034304"
/cranelation="NRTLPASGTDVAACQFLNALLRETKDMRYLPATVADLSOHI
PLSTQAIRVPRVPSPOHNOHNPAPMLIORSDSDAVTQTLIDLKPSYKGS
LDTETARFKQALBSHATWQALIDLRHMANIDRPLAFQAEOLLVGHAFHPAPK
SHPEPSESARRYLPDPASRPPLMPFAVRELIAGSLVSLERLIRAPASAPBL
GHFTDCMLPMPMOAHLQAMCQRLVANGDLDGAEAGOMLPTSSRSLSYSE
NSDMIKFSLVRLTNSVRLTSYKVGKRLARLAKTARMTLOARVPLRVQSEGM
AGLRDNGVIOESLALRVNLLFETPDSQTVIVSLQAPDGGDSILVAVRSLR
RDLPLAARACMVQAVCDRVLLPLPSAADGVILTAQCNLVEMODPVGILYR
DOCGMATGADBMKEAGEHVDENRFGSOLIRTPYLLNSTLAVTRALLAAGFD
SESLAQVRDALAEIRKTAQKTRCIDVLDSPWCKNPFICYLDRNENTADPAV
IYFDSNPYKKA"
13528..14475
/gene="lucB"
13528..14475
/gene="lucB"
/notes="Involved in aerobactin synthesis"
/codon_start=1
/cranet_table=11
/product="lucB"

Alignment Scores:
Pred. No.: 6.67e-25 Length: 36688
Score: 481.50 Matches: 112
Percent Similarity: 51.18% Conservative: 62.
Best Local Similarity: 32.94% Mismatches: 149
Query Match: 23.52% Indels: 17
DB: 1 Gaps: 6

US-09-889-746-2 (1-400) x AY046057 (1-36688)

QY 1 MetArgHisSerHisTyrPheGlnTrpLeuSerLeuProLeuSerValAlaValThr 20
Db 17563 ATGAAACGTTCTACTCTT-----TGGGTTTAAATCTTGCTTGCAATGCTTCT 17616

QY 21 GlnGlnLeuTyrAlaGlnProAsnGlnSerLeuProThrValGlnLeuGluProVal 40
Db 17617 ACCCTGCGTGGGGGGAAGAAAGAAAGAA-----GAAATATTGTT 17658

QY 41 IleThrIleAspIysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThr 60
Db 17659 GTCCTCCCGCCGACCGCGACCGCGCTGCG-----GAGATGCGCGCAACT 17706

QY 61 ThrIleValIleIleTyrGlnGlnGlnIleGlnGlnGlnAlaThrGlySerArgGlnLeuAla 80
Db 17707 ACCCTGGGTATTTGACCTGCTGCTGATTCAGGCTTCAAGTGAAGAGATTAA 17766

QY 81 AspValMetAlaGlnLeuIleProSerLeuGlyValSerSerGlyThrThrSerAsnPhe 100

```

```

FEATURES
    source
        University of Texas, Austin, TX 78712-1095, USA
        Location/Qualifiers
            1..20962
                /organism="Shigella boydii"
                /strain="0-1392"
                /db_xref="taxon:621"
                /note="aerobactin island SHI-3"
            84..1345
                /gene="int3"
                /note="similar to P4 integrase"
                /pseudo
            1800..2249
                /gene="shib"
            1800..2249
                /gene="shib"
                /note="similar to shib of Shigella flexneri M90T SHI-2
                presented in GenBank Accession Number AF141323"
            /codon_start=1
            /transl_table=11
            /product="Shib"
            /protein_id="AAK71629.1"
            /db_xref="GI:14626623"
            /translation="MDENALGTSYWRNSLADAESGSGSPERKAKNFTWHGIAAGR
            LDEAIVKFPKGGKDDVETDVILRPKYVFLRLQHGKDRSAGADIVTPIVTPALLSR
            EGFYPTPATSIPIRDLLEPLPKGAFSGEIGQYDKIKTHITFLYQL"
            3198..4459
                /note="similar to IS600 integrase; contains premature stop
                codons"
            4815..5081
                /gene="orf2"
                /note="similar to orf24 of Shigella flexneri SA100 SHI-2
                presented in GenBank Accession Number AF097520"
                /pseudo
            complement(5037..6229)
            /gene="orf3"
            /note="similar to orf25 of Shigella flexneri SA100 SHI-2
            presented in GenBank Accession Number AF097520"
            /pseudo
            5877..6305
                /note="orf4; similar to orf27 of Shigella flexneri SA100
                SHI-2 presented in GenBank Accession Number AF097520"
            /codon_start=1
            /transl_table=11
            /product="unknown"
            /protein_id="AAK71630.1"
            /db_xref="GI:14626623"
            /translation="WINTPSTAVVMALVAISTLPSRSRVKMLPVPRAHNTTGLLPV
            REICPHHGDDGRNSIERISRAAHTDLRFVCMTRGSGTTSRPFCTIPRSPALNAIG
            QDSGFVGVSGIPGDIIMFQHLVLIFFICKINLSDNNIYC"
            6308..8089
                /gene="iucA"
            6308..8089
                /gene="iucA"
            /codon_start=1
            /transl_table=11
            /product="iucA"
            /protein_id="AAK71631.1"
            /db_xref="GI:14626624"
            /translation="WR1IIDIIVILLCGSLTWTLPSEKPAIDVAQCFLNALIRET
            TDWKUTEYPPDELLIPLEQKSLHPRVAYFPTQHRPAPPARLVTAGSGSYPDVFTTL
            SLLIDKHLHQFLFUPVUCETPHORVLESHVHTQOAIIDARHDWAALREKALFNAGEAQ
            ALLTHAPHPAPKPEPNRREAREYLPDMAPHLPLRFWSVDKTIAGESLHLNQQR
            LTRAEANAPOLLNSQNLWFLPHWQGEYLQQGCOALVAKGLIKDIGEAGTSV
            LPTTSRSLYCATSMDMKFSLVRLTNSIRTLVSKVKRGWRLARLAOTDGMWLOV
            RFPFRWQEDGWLGLDNLGNIMQESIPALRENLVDQPSQTNVLSVLQAAPDGG
            DSLVSAVKRLDRLGIIVQQAHAHWDAYCQOVLKPLPTAEADYGLVLHAQCNILV
            QMGLDVGFIYRDCQGAFTROTGSILLARRRKIFSPMSCCALSLITCTPL
            FAVTALGAAGLDESINLMARVASLAEVRDQVTHKTCNLNVLESPPYNNVKNFFCYL
            NDHNENTIVDPSVIYFDPAFLQAEV"
            8090..9037
                /gene="iucB"
            8090..9037
                /gene="iucB"
                /note="similar to Shigella sonnei IS600"
                complement(15804..16160)
                repeat_region
                repeat_region

```

```

/codon_start=1
/transl_table=11
/product="iucB"
/protein_id="AAK71632.1"
/db_xref="GI:14626625"
/translation="MSEANIHSRYGLRCEKLPKLNGLWGLDNSAVLHCPGELPTCG
LCDALDOIPIAAPQLSNAVALPMAEWREEPQALTLFGQVKSIDIHRTAFWQLPLWLSPP
ANRASGEMVFADEREIYFPORPPRQGEVYRRYDPRIRRLMSRIADIPDAERFTRW
MNDPRVEYFWEQSGSLEVTAYLERQLTGKFAFPLIGCFDPRFPSPFEIYAAADRIG
RHSWOPFDRGLHLLVGEQOMGAHYVQSWLRLGTHYLLLDLDEPRTOTRTVLEPRTDNQR
LFRHLEPAGVRLTIKEPDPFKHRSVMWADRRHFFTEVGL"
9037..10779
/gene="iucC"
9037..10779
/gene="iucC"
/codon_start=1
/transl_table=11
/product="iucC"
/protein_id="AAK71633.1"
/db_xref="GI:14626626"
/translation="MNHKWDVFNRRRLVAKMLSEMEYEYQVFAESQGGDDHYCINLPGA
QWPIAERGIWGLWIDAQTLRCTDEPVLACTLLMOLKPVLSMSDATVAHMOPLYAT
LIGDLQLKARRGLSASDLNLADRLQCLLSGHPKFPVKNKGRGKGEALERYAPEY
TNTFRLHLAVKREHMIWRCNDLDDIQQLLTAAMDPOEFTFRSQVWQENGLDHNWLEL
FVHPWQQAIDTADFAEGRWVSLGEFGDQWLAAQOSLRTLTNARRGLDITKJPL
TYNTSCYRIGIPRYTAAGPLASRWLQQVFATDALTVOGAMILGEPAAGVSHEGYA
ALARAPYRQKMLGVIWRENPCRWLKPDESPLVLMATMECDENNQPLAGAYIDRSGLD
AETWLTOLFRVVVPLVHLLCRYGVALIAHQNTILAMKEGVORVLLKNVQGBMRV
KEAPPEDMSLPQEVRODVTSLRSADYLHDIQTGHFVTLRFISPLVRLKGVBERFQ
LLAAVLSYNNKHPQMAERFALFSLRPQIRVVVLPVLTWPDLOGGSRMLPNYLEN
LQNPMLVLTQYES"
10776..12113
/gene="iucD"
10776..12113
/gene="iucD"
/codon_start=1
/transl_table=11
/product="iucD"
/protein_id="AAK71634.1"
/db_xref="GI:14626627"
/translation="MKKSVDVGMVGTGPPNLSIALSHQIEELNCLFPDEHPHPSWHP
GMLVPDCHMTQVFLKDLVSAVATPNFYSLVLYVKKKFFRFLTSRLTYSREPSDY
LRWAEDMNNLYSHTVENIDFPKSRFLVLTSGRGEYFAHNICLGFKQKPLPFCVK
HVTQSCFHAEMLRRLPDLSGKRITVVGSGGADLFLNALRGSGAAEINWVSRRN
NFNALDEAFADYEYFTPEYISGFGLKEDIRHQLLDQKMTSDGITADSLITYRELY
HREVLKPRNIRLLPSRSVTLESSEPGWRLLMHHLDOGRESLESQVVIATGYRS
ALPOLPSLMPLITMLDKNTFKVRDDFTLEWSPKNNIFAVNASMOTHGIAEPQLSL
MARSARILNRVLGRDLFDLSMPALLQWRSRKKKQPEANSLTHYTAIOE"
12116..14314
/gene="iucA"
12116..14314
/gene="iucA"
/codon_start=1
/transl_table=11
/product="iucA"
/protein_id="AAK71635.1"
/db_xref="GI:14626628"
/translation="MMRKYMPRALGPLLVLVSPAVAOQNDNNEIIVCASRSNRTIA
EIAQTWVINEAELEQOICGKELKDALQALPGLDVSSOSRTNYGNMGRPLAVLI
DGVLRNRSRDSRQLDSVDPFNDIHIEVIGATALYGGSGTGLINIVTKGQDETMM
EFTAGKSGFNGSKDHERIAGAVSGNDHISGRLSVAYOKFGWFGDNGDALLDNT
QTGLQNSRLDIMGTLNIDEGRQLITQYYKQSGDDNYGLNLGSPASISGSTP
YVSKGLNSDRIPGTERHLISLQVSDSLFLQGLVGVQVYHRETSRLRYPPFPANAKQA
TAFSSQODTQYQMKLTLSQLMDGWOITWGLDAHERFTSNOMFLDQAASAGGL
NNKITYTGVPSVDITXLAFLQSSVDINDITVSGGVQVYQVTEVDFIDTQOQ
XIAGKALSADALPGCSVDNDFLNFAGLLMHITERQOAFNFSQGVALPDGKYGR
GTYGAAYVNSLTKSVNSDSKLEGVSDYSELGWRFTGNLRTQLAAYSLSNKS
ERNKDLTIVKDDRRRIYGVAGVDYLPDPTDSTGYNFNVLTKEVSGWQGVYDKV
ESSFSKATAYINWAPESWLRVQSTSTFSDAEGNDINGYTVDFISSWGLPVTLS
FSVENLFDROYTVMGQAPLYSPGYPSLVYDKGRGRTFGLNYSVLP"
14543..15810
/gene="iucB"
14543..15810
/gene="iucB"

```



```

repeat_region /note="partial copy of IS285 transposase"
               complement(16375..17107)
               /note="partial copy of IS200 transposase"
misc_feature 17707..17779
               /note="lex-like sequence"
gene          complement(17929..18591)
               /gene="L0006"
               /note="similar to Escherichia coli O157:H7 LEE prophage
               gene L0006"
               /pseudo
gene          complement(18605..18787)
               /gene="L0005"
               /note="similar to Escherichia coli O157:H7 LEE prophage
               gene L0005"
               /pseudo
gene          complement(18800..19048)
               /gene="L0004"
               /note="similar to Escherichia coli O157:H7 LEE prophage
               gene L0004"
               /pseudo
repeat_region complement(19700..20962)
               /note="similar to Shigella sonnei 18600"
BASE COUNT 5288 a 5258 c 5218 g 5197 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 1..666-24 Length: 20962
Score: 472.00 Matches: 133
Percent Similarity: 45.67% Conservative: 62
Best Local Similarity: 31.15% Mismatches: 170
Query Match: 23.06% Indels: 62
                Gaps: 11
US-09-889-746-2 (1-400) x AF335540 (1-20962)
QY 10 LeuSerLeuProLeuLeuSerValAlaValThrGlnGlnLeuTyAlaGlnProLeuGln 29
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12155 CTGCTTCTTCTGCTGCTGACCAAGCTGCTGCCAGCA-----AACGATGATTAATGAG 12208

QY 30 SerLeuProThrValGlnLeuGlnProValIleThrIleAspLysSerGlyMetAla 49
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12209 -----ATCATAGTGTGCTGCCAGCCGACGCAATCGAACT 12241

QY 50 LeuAlaAsnArgIleThrGlnMetProIleThrIleValIleTyGlnGlnIle 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12242 ATAGCG-----GAGATAGCGCAACCACTGGGTATCGAAATGCCGAATCG 12289

QY 70 GlnGlnGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSer 89
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12290 GAGACAGCAGATTTCAGGCGGTAAGAGCACTGAGCTCAGTTAATCCCGCGC 12349

QY 90 LeuGlyAlaSerSerGlyThrThrSerAsnProGlyGlnThrMetHisGlyArgGlnVal 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12350 CTTGATTCAGCAGCCAGACCAACCACTGATGAACATGCGTGGCCGCCCTG 12409

QY 110 GlnPheLeuLeuAsnGlyValProLeuThrGlySerArgAspIleSerArgGlnLeuAsn 129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12410 GTTGCTCTGATGACGGTTCGCTCACTTCACGTTCCGACAGCCCAACTGAC 12469

QY 130 SerIleAsnProAsnGlnValAlaArgIleGlyValLeuSerGlyAlaThrSerIleTy 149
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12470 TCTGTCAATCTTTAATATGACCAATATGAATGATCTCCGCGCGACGCTGTAC 12529

QY 150 GlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrIleYssAspLeuGlnGln 169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12530 GGTGGCGAGGTACCGAGGGTGTATCAACTCGTGACCAAAAAGTCAAGCCGGAACC 12589

QY 170 GlnPheGln-----ThrArgIleGlyValHisGlySerLysLeuSerSerGln 185
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12590 ATGATGAGATTGAGCGTGGACAAAAGTGGCTTTAAACGGCAATTAAGATCAGATGAG 12649

QY 186 GlyIleGlyTyGlnValGlyGlnSerValAlaGlyValSerGluAsn----- 203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 12650 -----CGCATTCGCGTGTCTCCGCGGAAATGACCAT 12685
QY 204 ValLeuAlaArgLeuAspValAspTyArgThrThrGlyGlyValaPheAspAlaAsnGly 223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12686 ATTCGAGACGCTTTTCCGCGGATATCAAAATTGGCGCGGTGTTGACGGTAACGGC 12745

QY 224 LysArgIleAlaProGlnProAlaGlnThrAspLysGlnAspSerLysSerLeuSerVal 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12746 GATGCCACCTGCTGTATAAACCCAGACCGGCTCAGACATCCATCGCTGACATC 12805

QY 244 AsnThrAsnValAspTrpGlnLeuAspAspLysGlnAsnIleAsnLeuAlaLeuThrHis 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12806 ATGGAAACCGTACGCTGAACATGATGATCCCGACGCTTCACTGATTAACGACGATC 12865

QY 264 TyrAsnAspLysGlnAspThrAspTyAlaProAspTyGlyAsnArgLeuAlaValLeu 283
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12866 TATTAAGTCAGGAGACACGACATTAACGGCTTAATCTCGGAAAGCTTTCCGCCATC 12925

QY 284 PheGlyGlyLys---ProSerLeuAsnAlaIleLysGlyLeuSerLeuSerGlnPro 302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12926 AGCGGAGACAGACACCATACGTCAGT-----AAGGGCTGATTTCTGACCGCATTC 12979

QY 303 LysThrThrLysSerThrPheAsnIleAsnTyHisHisAspAspLeuTrpGlyAsnThr 322
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 12980 GGCACGTAGCGGCAATTGATACAGCTGACGCTCTGACAGTATTTCTGGAGACAGAA 13039

QY 323 IleAsnThrAsnAlaTyTrpArgArgGlyLysGlyArgPheTyProPheValAlaPro 342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 13040 CTGCGCGTCAAGGTTTACTACCGCATGAGTCTTGCGGTACTACCGTTCCGACGGTA 13099

QY 343 PheSerIleAlaLysAlaLeuProIleLeuGlnSer----- 354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 13100 AATCGCAATTAACAGCGACGCGCTTCTCTCGTCACAGCAGATACCGACAGTACGGC 13159

QY 355 -----MetAsnLeuProSerAlaThrLeuAspAla----- 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 13160 ATGAACTGACTCTGAACAGCACTGATGAGCGGTGCCAATCACTGGGCGTGGAT 13219

QY 365 -----TyThrLysAlaProGlnAlaArgAla 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 13220 GCTGAGCAGTACGCGTTTACTTCAACCAAGATTTCTTCATGCTGCTCAGGCAAGTCT 13279

QY 374 TyrGlyValLeuGlnSerGlnSerLysValaGlnValLeuGlyArgValProAsnLeuAsn 393
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 13280 TCGGAGCGGCTGAACAACAT---AGATTACACCAACCGGCGGTATCGTATATGAC 13336

QY 394 LysProLysArgAlaLeuPhe 400
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 13337 ATCACCAATCTGCGCGCTTC 13357

RESULT 13
AF141323 23771 bp DNA linear BCT 20-JUL-1999
LOCUS Shigella flexneri SH1-2 pathogenicity island, complete sequence.
DEFINITION AF141323
ACCESSION AF141323.1 GI:5532445
VERSION Shigella flexneri.
KEYWORDS Shigella flexneri.
SOURCE Shigella flexneri.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
          Shigella.
REFERENCE 1 (bases 1 to 23771)
AUTHORS Mose, J.E., Cardozo, T.J., Zychlinsky, A. and Grosman, E.A.
TITLE The self-associated SH1-2 pathogenicity island of Shigella flexneri
JOURNAL Mol. Microbiol. 33 (1), 74-83 (1999)
MEDLINE 99340540
PUBMED 10411725
REFERENCE 2 (bases 1 to 23771)
AUTHORS Mose, J.E., Grosman, E.A. and Zychlinsky, A.
TITLE Direct Submision
JOURNAL Submitted (07-APR-1999) Microbiology, NYU Medical Center, 540 First
          Avenue, New York, NY 10016, USA
FEATURES
          Location/Qualifiers

```





**FEATURES**  
**source**  
**BASE COUNT** 803 a 859 c 828 g 733 t  
**ORIGIN**

**Alignment Scores:**  
 Pred. No.: 7,24e-25 Length: 3223  
 Score: 463.00 Matches: 134  
 Percent Similarity: 45.14% Conservative: 61  
 Best Local Similarity: 31.02% Mismatches: 177  
 Query Match: 22.62% Indels: 60  
 DB: 6 Gaps: 9

US-09-889-746-2 (1-400) x A07458 (1-3223)

QY 6 TyrPheGlnTrp-----LeuSerLeuProLeuLeuSerValAlaValThrGln 21  
 Db 383 TATACGCTTTGGCTCTCAACCCACTGCTCTTACCATGATGGCGGCAGCTCGTCAA 442  
 QY 22 GlnLeuTyrAlaGlnProAsnGlnSerLeuProThrValGluLeuGluProValValIle 41  
 Db 443 CAA----- 445  
 QY 42 ThrIleAspLysSerGlyMetAlaLeuAlaAsnArg-----IleThrGlnMet 57  
 Db 446 ACCGATGATGAACACTGCTGGTGTCTGCGCAACCGCAGCAATCGCACCGGAGATG 505  
 QY 58 ProHisThrThrLysValIleTyrGluGluGlnIleGlnGluAlaThrGlySerArg 77  
 Db 506 GCGCAACCACTGGGTTATCGAAACCGCGAACTGGAACAGCAGATTCCAGGCGGCANA 565  
 QY 78 GlnLeuAlaAspValMetAlaGlnIleProSerLeuGlyValSerSerGlyThrThr 97  
 Db 566 GAGCTTAAAGACGCACTGGCTCAGCTGATCCCTGGCTTGGAGCTCAGCAGCGGAGCCGC 625  
 QY 98 SerAsnPheGlyGlnThrMetHisGlyArgGlnValGlnPheLeuLeuAsnGlyValPro 117  
 Db 626 ACCAACTACGGTATGAATGTGGTGGCGCCGCTGGTGGTGGTGGTGGTGGTGGTGGT 585  
 QY 118 LeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAsnProAsnGlnValAla 137  
 Db 686 CTCAACTCTTCACGTCACGACAGCGCAGCAACTGGACTCTATAGATCCTTTAATATGCAC 745  
 QY 138 ArgIleGluValLeuSerGlyValThrSerIleTyrGlySerGlyValAlaThrGlyLeu 157  
 Db 746 CATATTGAATGATCTCCGGTGGAGTCTCCCTGACGGCGGCGGAGTACCGGTGGCCTG 805  
 QY 158 IleAsnIleValThrLysSerAspLeuGluGluGluGlnPheGluThrArgIleGlyVal 177  
 Db 806 ATCAACATCGTGACCAAAAGAGCGCGCAGCGGAAACCATGATGGAGTTTGGAGCTGGCACC 865  
 QY 178 HisGlySerLysLeuSerSerGluGlyIleGlyTyrGlnValGlyGlnSerValAlaGly 197  
 Db 866 AAAAGTGGCTTTAGCAGCAGTAAAGATCACGATGAACGATTTGGCGGAGCTGTCTCCGC 925  
 QY 198 ValSerGluAsnGlyAsnValLeuAlaArgLeuAspValAspTyrArgThrThrGlyGly 217  
 Db 926 GGAATGAG-----CATATCTCCGAGCTCTTCCGTGGCATATCAGAAATTTGGCGC 979  
 QY 218 AlaPheAspAlaAsnGlyLysArgIleAlaProGluProAlaGlnThrAspLysGlnAsp 237  
 Db 980 TGGTTTGAACGTAACGGGATGCCACCTTGTCTTGTATACACACAGCAGCGGCTGCAGTAC 1039  
 QY 238 SerLysSerLeuSerValAsnThrAsnValAspTyrGlnLeuAspAspLysGlnAsnIle 257  
 Db 1040 TCCGATCGCTGGACATCATGGGAATCGTACGCTGAACATCATGAATCCCGGAGCTT 1099  
 QY 258 AsnLeuAlaLeuThrHisTyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGly 277  
 Db 1100 CAGTTGATCAGACAGTACTATAAAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1159

QY	42	ThrIleApLySerSerglyMetAlaLeuAlaAspArg-----IleThrIlnet	57
Db	446	ACCCATGATGAAAGCTTCGCGGTGCTGTCTGCCAACCGACGAAATCCGACCGTACCGAAGTG	505
QY	58	ProHsIThrThrvValIleTyrgIugIuInIleInIugIuInAlaThrvIseArg	77
Db	506	GGCGAAACCACTGGGTTATTCGAAAAGCGGCAACTGGAAACAGATGTTACGGGCGGCAAA	565
QY	78	GlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSerSerglyThrThr	97
Db	566	GAGCTTAAAGCGACACTGGCTCAGCTGATCCCTGGGCTTGAAGTCAAGCAGCGAGCGGC	625
QY	98	SerAspHeGlyGlnThrMetHsIglTyArgGlnValGlnPheLeuLeuAsnGlyValPro	117
Db	626	ACCACTTACGGTATGAAATGTGCGTGGCGCGCCCTGAGTGTGCTGTGATTCAGCGCGCTGT	685
QY	118	LeuThrgIySerArgAspIleSerArgGlnLeuAsnSerIleAsnProAsnGlnValAla	137
Db	686	CTCAACTCTTACGTTACCGACGACGCGCAACTGAGCTCTATGATCTCTTAAATATGAC	745
QY	138	ArgIleGlyValLeuSerGlyValAlaThrSerIleTyrgIySerGlyValAlaThrGlyLeu	157
Db	746	CATTTGAAAGTAACTTCCTCGGTCGACGCTCCTGTACGGCGCGCGAGTACCGGTGGCTGT	805
QY	158	IleAsnIleValIleThrvSerSerAspLeuGlnIugIuInIugIuInPheGlnThrvArgIleGlyVal	177
Db	806	ATCAACATCTCGTGAACCAAAAAGCCAGCCGAAACCAATGAGATTTGAGGCTGGCACCC	865
QY	178	HsIglSeryIyLeuSerSergIugIyIleGlyTyrgIlnValGlyInSerValAlaGly	197
Db	866	AAAAGTGGCTTTACACACAGTAAAGATCAAGATGAAGAACGATTCGCGAGCTCTTCGGC	925
QY	198	ValSergIuAsnGlyAsnValLeuAlaArgLeuAspValAspTyArgThrvThrGlyGly	217
Db	926	GGAAATGAG-----CATATCTCCGACGCTCTTCCGTGCATATCAAGAAATTTGGCGGC	979
QY	218	AlaPheAspAlaAsnGlyValArgIleAlaProGlnProAlaGlnThrvAspLeuGlnAsp	237
Db	980	TGGTTTGAAGGTAAACGGCGAGTACCACTTCCTGTTGATACCCAGACGGGCTGTGAGTAC	1039
QY	238	SerIySerIeuserValAsnThrvAsnValAspTyrgIlnLeuAspAspLeuGlnAsnIle	257
Db	1040	TCCGATCGGCTCGACATCTATCGGAACCTGTAACGCTGAACATGATTAATCCCGGACGCTT	1099
QY	258	AsnLeuAlaLeuThrvHsItyrAsnAspIyGlnAspThrvAspTyrgAlaProAspTyrgIy	277
Db	1100	CAGTTGATCACACAGTACTATAAAGCAGAGGAGAGACGATTAACGGCTTAATCTCGGG	1155
QY	278	AsnArgLeuAlaValLeuPheGlyGlyIyProSerIeuserLeuAsnAlaIleIySgIyLeuser	297
Db	1160	AAAGCTTCTCGCATCAGAGGACGAGACGACGCACTTC--GTCAGTAAACGGGCTGAT	1216
QY	298	LeuserGluGlnIProIySthrvThrvIySerThrvPheAsnIleAspTyrgHsIAspAsp	317
Db	1217	TCCGACCGTATTCGCGGACCTAGACGGGCAATTGATCAGCTTCAGTACTCTGACAGCGCT	1276
QY	318	LeuTyrgIyAsnThrvIleAsnThrvAsnAlaTyrgTyrgArgIyGlyIySgIyArgPheTyrg	337
Db	1277	TTTCTGGGACAGAGAGCTGTGCGTCAAGTTTACTACCGGAGATGAGTCTGTGGGATTTTAC	1333
QY	338	ProPhe-----ValAlaProPheSerIleAlaIyAlaIeu	349
Db	1337	CGGTTCCGACGAGTAATTCGAAACAAACAGATGACGGCTTCTTCCTTCACAGACAGGAC	1396
QY	350	ProIleLeuGlnSerMetAsnLeuProSerAlaThr-----	361
Db	1397	ACCGACCACTACGGCATTAACCTGACTCTGAACAGCAAAACCATGACGAGCTGGCAATC	1456
QY	362	-----LeuAspAla-----	368
Db	1457	ACCTGGGGGCGTGAATGCTGATCATGACGCGTTTAACTCCCAACAGATGTTCTTTCACGCTG	1516
QY	369	ProGlnAlaArgAlaTyrgIyValLeuGlnSerIeuserIyAlaGlnValLeuGlyArg	388

Db	1517	GCTCAGGCAAGCGCTCCGGAGGGGTGAACAAGA---AGATTTCACCAACGGGGGC	1573
Oy	389	ValProaenLeuAenLysProLyAeRgAlaLeuPhe	400
Db	1574	TATCCGTCGTATGACATCACCAACTCGCGGCTTC	1609
RESULT 16			
LOCUS	ECDPIJ3RE	2769 bp	DNA linear BCT 24-FEB-1992
DEFINITION	E. coli plasmid pF88 cloacin DF13/aerobactin receptor gene.		
VERSION	X05874		
KEYWORDS	cloacin DF13/aerobactin receptor.		
SOURCE	Escherichia coli.		
ORGANISM	Escherichia coli Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.		
REFERENCE	1 (bases 1 to 2768)		
AUTHORS	Krone,W.J.A., Stegehuis,F., Koningsstein,G., van Doorn,C., Roosendaal,B., de Graaf,F.K. and Oudega,B. Characterization of the PCoIV-K30 encoded cloacin DF13/aerobactin outer membrane receptor protein of Escherichia coli; isolation and purification of the protein and analysis of its nucleotide sequence and primary structure FEMS Microbiol. Lett. 26, 153-161 (1987) 2 (bases 1 to 2769)		
JOURNAL	Oudega,B.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (24-FEB-1992) B. Oudega, Faculty of Biology, Universiteit de Amsterdam, De Boelelaan, 1081 HV Amsterdam, Netherlands		
TITLE	Data kindly reviewed (21-JUL-1988) by Oudega B.		
JOURNAL	Location/Qualifiers		
COMMENT	1. .2769		
FEATURES	/organism="Escherichia coli"		
SOURCE	/strain="F344"		
	/db_xref="taxon:562"		
	/clone="PCoIV-K30"		
	/plasmid="pF88"		
RBS	157. .164		
CDS	159. .2367		
	/codon_start=1		
	/transl_table=1		
	/product="receptor protein precursor"		
	/protein_id="CAA29297.1"		
	/db_xref="GI:41262"		
	/db_xref="SMISS-PROT:P14542"		
	/translation="MMISKYYTLMAALPLITMMAAPAQAOTDDETFPVNSNRRTYV AEAMQTTVINALEBOOIQQGKKLKALAAOLIGLDVSSSRNTYGNVRGRPLVALI VDGRLLNSRTDSRLQSDIDEPNNMHLEVIFGATSLVGSGSTGLINIVTKGGPETT MEFPAGTKSGPSSSKXDERLAGAVSGNHEISQLSVAYOKFGWPGKNDARLLDNL TONTLOYASRLDIGTGLTNLIDESRQOLIROYKSGODDDYGILNGSPAIIGTSIS PTONGLNISDRIPETGDHLISLOYSDAFLGOBLVGVOYVEDSLRFPPVTNANKCP VTAASSQOOPDJOCMTKLANSKRMDMOITWGDLADHERPTSOMEPDLAASAAGC LNNKAITTGTRIPESYDITNLNAFLQSGLDINNLTTLNGCARVQTENKIDPFIYAQC ROIAGAKATSADAFAWRLSRLHFLFNAGLIMHITEPOQAWLNFSGLELPBGKYRGGA GIVAAVANGLPLTKSVNVSDSKLEGVDSYEELGMFTGNMLTOIAAYSIDKSYSIDK VANDOLISVVDDRRIYGVGAVDYLIPDTMSGTGFVNVLKTESKVGNTWKQYDVN TASASKATAYIGMAPDPMSLRVOSTGSFDVSADAGYKVDTGYTDLGSYOLPGTLLS PSIENTLDRDRTTYMGORAPLYTPGIGPASLYYKRGRTFGNLYSLP"		
sig_peptide	169. .243		
CDS	244. .2367		
	/note="Protein sequence is in conflict with the conceptual translation"		
	/codon_start=1		
	/transl_table=1		
	/product="mature receptor protein"		
	/protein_id="CAA29298.1"		
	/db_xref="GI:1333741"		
	/db_xref="SMISS-PROT:P14542"		
	/translation="QOOTDEPFVNSNRNTVAEMQTTVINALEBOIQGKEIK KDAAQLRIGLDVSSSRRTYGNVRGRPLVVDGRLLNSRTDSRLQSDIDEPNNH HIEVIPATSLVGGSGTGLINIVTKKGPEPTMNEFELGTSGSSSDKHERRLAGAA		



**TITLE** Barrio, Huber, F., Bowser, L., Capela, D., Galiberti, F., Gonzy, J.,  
Gurtal, M., Hong, A., Hutzar, L., Hyman, R., Kahn, D., Kahn, M.,  
Kallman, S., Keating, D. H., Palm, C., Peck, M. C., Surzycki, R.,  
Wells, D. H., Yeh, R.-C., Davis, R. W., Surzycki, N. A. and Long, S. R.  
**Direct Submission**  
**Submitted** (29-MAR-2001) Biological sciences, Stanford University,  
371 Serra Mall, Stanford, CA 94305, USA  
**Location/Qualifiers**

**FEATURES**

**SOURCE**

```

/organism="Sinorhizobium meliloti"
/strain="1021"
/db_xref="taxon:382"
/plasmid="pSymA"
115. .235
/gene="ntrbA"
/note="SMA2414"
115. .235

```

/t/raue/ati/on="MANNENGIGISCERVAVVIGFETGVAOEAPENOSATVLSLEIVV  
 TGRRAAOOISLEIARTIYYVDSOTIOAEKSCVGTLOOLIGETIPESDPADARSATVSG  
 NIKRGPDLPIVAVGSMVSARLSROFDLIPENITRVTVLSGATVYGNATGGIININ  
 ITRKKAEPLVGLAAEVTCMGSGFSGPODEPBNRANAGATVSNEMWDLIATNGRTVA  
 FIDGSGTLLIPIDITOTSTAFNERIDIMSGISTGQIDDBRAREVSCGYFBSKODSYGLY  
 YGFPAALADBSLPEFETRGSYEDENPÖTRKRSALVATYDNOVFCQGLDQMSYTERIN  
 KHPAPASNSEGTGPYFYGSSODTIDYVIRALVAEPTDAKIYVQDILDAMDSTPAQ  
 NIPDVAVAGSGSGDLFNTIGKTGLPSYDIDSVNGVAFASAEATDRILINGGVAVYCF  
 NTEVSDIFGAQGOVALLQORATSADTIPGGEVNAVDAALFASATYOLTNTOOVVANS  
 OGEFLDPKPKYXGIGNVSFSGGHYTLVANSVAVGEGKAVLAKTNSGEITRELDCTFNNI  
 EBAAYVLSDBREININRSALVEIIDREVRVYJGEGKAVLADHCPGVGVGHVYREDE  
 YKAGDGEWEDSVGSASVSKTSGYVGMFTDALSIFRSGCHITBELDQNFITLDDYTDLE  
 LTGGRFRENFTDITLNPGLHNVFDTDITYTVMGSRAKALVGLADESVFDPYKRGKRTFVAV  
 SLTKXF"

gene	2355. .3467
	/gene="Sma2339"
CDS	2355. .3467

```

/function="Sma2339"
/name="Transp"
/notes="glimmer prediction; similar to iucB, aerobactin siderophore biosynthesis protein from E. coli, Q47317, a from Shigella flexneri, AAD44747"
/codon_start=1

```

```

/translation table=11
/product="putative siderophore biosynthesis protein"
/protein_id="AAK65924.1"
/db_xref="GI:14524438"
/translation="MRGRGCGMGSGRAGGGVAGPEVDQADGAPFSLHAGCGVELEY
EASGILASASGRRAPAPAGGGRIDRLSCNAGLALCAAPYLAPATPSSVLYE
LAEGCNVUSPCKTCRGILVFNATLTHSTTFAEMFQVDEYMAASPSVTPPRDHPD
GRREHLPSPKPKAGCTYAPITPGLSTLSLAFATLTLNDEIDHRMKNRNVAEENKAG
SKAGAGYERKNAADHTILPLIGRFNAPAFASIFELIYMAKEDVIGPSAGGIDYDCK
IVAESCRGQVMTSLPDLMLHMTADDRTIRIVQEPASAHHRLQGLQSSGFSHTIR
VLDLPTKRAATMSIASHQRFPEPNTLMHLPDDPRNS"

```

gene	complement (36668 .	.4543
CDS	/gene="Sma2341"	
	complement (36668 .	.4543)

```

/gene="Sma2341"
/function="Miscellaneous; Not classified regulator"
/note="glimmer prediction; similar to several putative
transcriptional regulators, including Xylella fastidiosa
AT884577; E. coli, P75836 and Pseudomonas aeruginosa,
010610. Helix-turn-helix motif"

```

gene  
CDS

/product="putative LysR-family transcriptional regulator"  
/protein\_id="AAK65925.1"  
/db\_xref="GI:14524439"  
/translation="MNLLPLFLVAAEEDNPPAADRLGYTRSAVSQGIIRLEDAFGTT  
LVTTRSVRLTEAGERLREALSQPLSDIGALTRVAGEBDEPRGLLIIVATSIAGEFL  
SGPILVAPAAQHPFVALIDVTYDDEFPDIVAGFPAGVAGLVEIQDNIAPVLTSDQER  
LAVLAPLTAARGTTPADHPRLVMHRCIGMRVAPNVAVYRMFENGGVDFVDVAEPQITL  
TNDRLFMRLSLCGAGITFAATEETFRPFVEGELVPLDQLPPGPFVFLYFPQRUMNR  
TPKRRALLVHTRRRR"  
4655. 5377  
/gene="SM2343"  
4655. 5377

/gene="Sma2343"  
/function="Small molecule metabolism"  
/note="glimer prediction; similar to putative  
oxidoreductases from Streptomyces coelicolor, CAB59579 and  
CAB58279, as well as to a hypochemical protein from  
Synecocystis, S76993; short-chain family"

```

/gcdon_start=1
/transl_table=1
/product="putative oxidoreductase"
/protein_id="AAK65926.1"
/db_xref="GI:14524440"
/translation="MDKVIITLGASGGIGESIAELGVAGAKILLGARQARIEAIAT
EIRBAGALAOVLDTDRHSVAFAOAAVDTWRIDVLVNNAGVMPISPLAAKVEDE
MERADVNIKVYLGNGICAVLEPIMAEQSGQIINGISGALSVPFAAYCTKTRAYRA
ISDGRQDSINIRIVCPNPGVSESLGTTIHEBTAMMDTYRAIALQPAIDIAVRGQ
VIEAPQSDVSTIRITRCPASGN"
/complement (5444 . 6016)
/gene="SMA2345"
/cds
/complement (5444 . 6016)

```

```

/gene="SMA2345"
/function="Miscellaneous"
/note="glimmer prediction"

```

```

/codon_start=1
/transl_table=1

```

```

gene
CDS
        /product="hypothetical protein"/
        /protein_id="AAK65927.1"/
        /db_xref="GI:14524441"/
        /translation="MDVYAAAGHPSGPAORSLGTRILVYVNDVRVIGRTETKSGFLD
        HRSVTHOYTPDELSIVRSRSRGCRSVEPDDDRISIGEDMNAHPNHQCEIADGSGSKG
        PLDELIDTVASGSGSTVDCPDIDMRKLCQDVAPCSPFLECAHSGSILRLDIGMPP
        DGIIGRSRNSAHSRRSDDOHADAAAGALWQ"
        complement(6391..7128)
        /gene="SMA2347"/
        complement(6391..7128)

```

```

/gene="SMa2347"
/function="Miscellaneous; Hypothetical/Global homology"
/note="Glimmer prediction; similar to Bacillus halodurans
ORF, BAa57318"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK65928.1"
/db_xref="GI:14524442"
/translation="MTKASCNIVCYWSESDRMRPSPFEGAPFAAASHVYSTGS
KVPGLAALHWTSMVLAERAVQARVIVTGAAGMTEKALADENSWSFGCISDAPM
LRVAEQTGPHILRVHNDQGITGAAPSPFPANVCLILTVHVGAKRDLDTIEQIRRL
VYGAQFVAHVSIPOSEPERSTWIAIRHVAFGTSGSEAESARQAIATKLSVLSPEDE
AVLRAGSGSDVRLPFAAMTFRGWGYA"

```

gene	7431. .8078
CDS	/gene="Sma2349" 7431. .8078

```

/gene="SMA2349"
/function="Small molecule metabolism"
/notes="Glimmer prediction, very similar to putative
oxidoreductase from E. coli, P7165, and from Deinococcus
radiodurans, B75576. 2Fe-2S binding domain. [2 Fe-2S]
binding domain. May be one of three subunits of
(SMA2349, SMA2351, SMA2353)"
/codon_start=1
/trans_table=11
/product="probable oxidoreductase"

```

```

/protein_id="AAK65929.1"
/db_xref="GI:14524443"
/translation="MELDMQSPGALERTSRRLDAAATVTVSAHSLGHQAOTNASTA
PQSTKVTPTVNDERELQDNRTLLDALREHLHJTGKGCDDHGCCACTVLIEGRR
VACLTLAIHEGDSITTTLEGLOPENLHPMOAFVKHGDGFCQGYCTPGQICSSVAML
DEIKANIPSHVTDLTAPAEITPAIRERMSGNICRCGAYSINVEAITEVAGRA"
8075..9025
/gene="SMA2351"
CDS
8075..9025
/gene="SMA2351"
/function="Small molecule metabolism"
/note="glimmer prediction: very similar to E. coli YagS, a
hypothetical protein, P77324; may be one of three subunits
of (SMA2349, SMA2351, SMA2353)"
/codon_start=1
/translate=1
/product="possible oxidoreductase, molybdopterin-binding
subunit"
/protein_id="AAK65930.1"
/db_xref="GI:14524444"
/translation="MRAFSYERSVSGIAAARAAAATDGAKFIAGTGNLLDLMLKEIET
LRNLVDNGLDRIETAREGGLRIGALVRNTDLAANDRERDVALLSRLLTGASGO
PHTATTAGNLLORTRCPFYDTNPKRPGSGCSALAGFSRLQAVVGVSDACIAS
HPSDMAVALRALDAVETVRADGTTTSPIMGDLHRLPGDTPHIEHVLBERGLITAVVL
PKPAGGKQIYKRYDRASYPALVSVAVOPDGTGRVAVGVGIAPKPRDRDAEKELA
NGAKAVAAMLLAGARPTQONAFKTLVERALSVLIVEAK"
9027..11228
/gene="SMA2353"
CDS
9027..11228
/gene="SMA2353"
/function="Small molecule metabolism"

```

## Alignment Scores:

```

Pred. No.: 1.93e-23 Length: 11334
Score: 453.00 Matches: 109
Percent Similarity: 52.69% Conservative: 67
Best Local Similarity: 32.63% Mismatches: 132
Query Match: 22.13% Indels: 26
DB: 1 Gaps: 9

```

US-09-889-746-2 (1-400) x AE007312 (1-11334)

```

Qy 18 AlaValThrGlnGlnLeuTyrrAlaGlnProAenGluSerLeuProThrValGluLeuGlu 37
Db 184 GCTGTTGCCAGGAG-----CCAGCGAATCAATCCGAGCTGTGACGAGCCTGGAA 234
Qy 38 ProValValIleThrIleAspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMet 57
Db 235 GAAATCGTAGTCACCGCGGAGCATCC-----GCTCAGCAGATTTCTGAAATC 282
Qy 58 ProHisThrThrLysValIleTyrrGluGluGlnIleGlnGluAlaThrGlySerArg 77
Db 283 GCGCGCACCATTTACGTCGTCGATTCGGATCGATCAGATCCAGCGCGCGGTGCGGCAAG 342
Qy 78 GlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGly---ValSerSerGlyThr 96
Db 343 ACGCTGCAGCAGATTCTGGCGCAACCATTTCCGAGCTTCGACCCGCCAGCAGCGAGCG 402
Qy 97 ThrSerAsnPheGlyGlnThrMetHisGlyArgGlnValGlnPheLeuLeuAsnGlyVal 116
Db 403 CGTAGCTCATTCGGCCAGAACTCGTGGCCCGGCCACCACTGATCTTGTCCAGCGTGTC 462
Qy 117 ProLeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAsnProAenGlnVal 136
Db 463 TCGATGAACCTCGCGCGCTCGTGCAGCGACCAATTCGATGCCATCGATCCATTCAACATC 522
Qy 137 AlaArgIleGluValLeuSerGlyAlaThrSerIleTyrrGlySerGlyAlaThrGlyGly 156
Db 523 GAACGGGTGGAGGTCTCTGTCTGGGCAACTGCAATCTACGGCGGGAACCAACCGCGCGC 582
Qy 157 LeuIleAsnIleValThrLysSerAspLeuGluGluGlnPheGluThrArgIleGly 176
Db 583 ATCATCAACATCATCACGAGAGGACGAGGACCGCGGACCG-----GGC 627

```

```

Qy 177 ValHisGlySerLysLeuSerSerGluGlyIleGlyTyrrGlnValGlyGln----- 193
Db 628 CTGCATTCGGGAAGTGACCGCGCGCATGGCGCGCTTTGCGCGCAGCGAGCTTCGAC 687
Qy 194 ---SerValAlaGlyVal-----SerGluAsnGlyAsnValLeuAlaGlyLeu 208
Db 688 CGCAATCGCGCGGTGCGGTCACTATTAACAGCGAAACTGGGAT-----GCCCGCCTT 741
Qy 209 AspValAspTyrrArgThrThrGlyAlaPheAspAlaAsnGlyLysArgIleAlaPro 228
Db 742 TCGATCGCGCGCAACAGACCGCGCTTTTATGACGCGCGGCGACGCTGCTGATCCCC 801
Qy 229 GluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsnThrAsnValAsp 248
Db 802 GACATCCAGCAGAGCTCCACCGCATTCACGAGCGCATCGACCTGATGGGTCTATCGGC 861
Qy 249 TrpGlnLeuAspAspLysGlnAsnIleAsnLeuAlaLeuThrHisTyrrAsnAspLysGln 268
Db 862 TACCAGATCCGACGACGCGGTGCGGTGCAATTTCCGGGCAATATTTCCGACGCAAGCAG 921
Qy 269 AspThrAspTyrrAlaProAspTyrrGlyAsnArgLeuAlaValLeuPheGlyGluLysPro 288
Db 922 GACTCGGATTACGGGCTTTATTCGGGCGCTTCTTTGACGCGCTC-----GCCGACCG 975
Qy 289 SerLeu---AsnAlaIleLysGlyLeuSerLeuSerGluGlnProLysThrThrLysSer 307
Db 976 AGCCTGTTGAAACCGGTTCCGGATACGAGTCCGACTTCAATCCGACACACGCGCTCG 1035
Qy 308 ThrPheAsnIleAsnTyrrHisAspAspLeuTrpGlyAsnThrIleAsnThrAsnAla 327
Db 1036 ATGTTGAACGTCACCTATACCGATAATGACGTTTTCGTCAGCAGCTATTGCTTCAGGGA 1095
Qy 328 TyrrArgArgGlyLysGlyArgPheTyrrProPheValAla 341
Db 1096 TCTTACCGAGCTGAGCGGATAAGTTCCATCCCTCCCTGCT 1137
RESULT 18
AF110737
LOCUS
DEFINITION Sinorhizobium meliloti strain 2011 rhizobactin regulon, complete
sequence.
ACCESSION AF110737
VERSION AF110737.1 GI:4151930
KEYWORDS
SOURCE Sinorhizobium meliloti.
ORGANISM Sinorhizobium meliloti.
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Sinorhizobium.
REFERENCE 1 (bases 1 to 11853)
AUTHORS Lynch,D., O'Brien,J., Welch,T., Clarke,P., Cuiv,P.O., Crossa,J.H.
and O'Connell,M.
Genetic organization of the region encoding regulation,
biosynthesis, and transport of rhizobactin 1021, a siderophore
produced by Sinorhizobium meliloti
J. Bacteriol. 183 (8), 2576-2585 (2001)
21172875
PUBMED 11274118
REFERENCE 2 (bases 1 to 11853)
AUTHORS Lynch,D., O'Connell,M. and O'Brien,J.
TITLE Cloning and sequence analysis of the Sinorhizobium meliloti 2011
rhizobactin regulon
Unpublished
REFERENCE 3 (bases 1 to 11853)
AUTHORS Lynch,D., O'Connell,M. and O'Brien,J.
TITLE Direct Submission
JOURNAL Submitted (03-DEC-1998) School of Biotechnology, Dublin City
University, Glasnevin, Dublin, Leinster Dublin 9, Ireland
FEATURES
source
1. 11853
/organism="Sinorhizobium meliloti"
/strain="2011"
/db_xref="taxon:382"

```



```

gene      8. .11848
           /gene="rhizobactin regulon"
misc_feature 8. .26
           /gene="rhizobactin regulon"
           /note="FUR Binding Site"
gene      42. .1454
           /gene="rhba"
           /note="previously rhba"
           /gene="rhba"
           /function="siderophore biosynthesis"
           /note="similar to DABA aminotransferase"
           /codon_start=1
           /transl_table=11
           /product="Rhba"
           /protein_id="AAD09412.1"
           /db_xref="GI:4151931"
           /translation="MPADLAARTSSKTFNGVDLMDASAPADNAFYIDROERESNARS
YPRRPVVALKSGSCIVTDGRSYLDGAGTGLAGNHPEVLEITLQQVAGSGPL
HTLIDLTPVDRFVSDIFGTLPAGLDEAKIQCPSPGTDVAEALIKLAKTATGRTL
VSEFGAYHGMGSGSLMGLSGPKASVGQLVGAPFPYPYAVRCPFGSGNETATLA
AYEPERALDPGGINRPAVALLEAVQSGGVIPAVEWLRAVRVTRDLGIPLVDE
VOSGVRGTSFYAFQKAGIIPVYVUSKAIIGGGLP LAVITYREDLDMKPGMAHGR
GQOLAPAGSKTLEITERELVERRAIAGRRLANLERIAQTPYIGVYRGGLMGV
EYVDPGLPDALGHPHGOEIARMIOHEMFRAGITLLETGRGVSVALPLPVIISDAE
IQVSGALAAAEERLGRKAA"
gene      1451. .2938
           /gene="rhbb"
           /note="previously rhbb"
           /note="previously rhbb"
           /function="siderophore biosynthesis"
           /note="similar to DABA decarboxylase"
           /codon_start=1
           /transl_table=11
           /product="Rhbb"
           /protein_id="AAD09413.1"
           /db_xref="GI:4151932"
           /translation="MNINVAAPFTPTKOTDHAQDILGTSSESRVFRNMLQALDMV
VNOTAASSLVSSTGTSFQGLRGLIDLDLPEVCTGIAALAEGRPALSHANVYGRPA
GMATHCFVAVPALAEVILISATNOSLSDMSDPSFATLVEERYLACTLQALPMSAS
GNFTSGTQSNMTALYLAARCGPDARKAGVATLVEERYSIKSNAIIGFADAVIA
IADADGRMSVPALKAEILRVAGEGRI PVAVATAGTDLGADPLVEIADIAAQN
MHWDAAYGGGLFSRHSRLBGLBEAHSITLDFHMLFOPISCGVLLRDADEFPL
ASKADYINPEDAVFADAPNLVRSMTERRADAKILMTMBRAGIDGGLALICOTLON
TAAARAAVTRREYLSAGPSLSTVFRVVSARGPPPAATILKTPAALPMGIALA
TIVLDGRVHFKITLLNPKSTPDVVRILDAIGTAELESTHARP"
gene      2925. .4682
           /gene="rhbc"
           /note="previously rhbc"
           /note="previously rhbc"
           /gene="rhbc"
           /function="siderophore biosynthesis"
           /note="similar to Iuca"
           /codon_start=1
           /transl_table=11
           /product="Rhbc"
           /protein_id="AAD09414.1"
           /db_xref="GI:4151933"
           /translation="MHAHNAATLTLRSLNCVAREPDPHVMLETQHLRFMLTPREG
GSLGLPAHRSATGHHLFGEPVMTDENCAKVDVAEIASVIERLEPSIAKDRV
DLNLRTHSSRLILBALHARNGDLALAGDEVSVAAEQGLIAGHIGHPCPSREGMT
EASRRYSPEFAAGPLRWAESSELPHGHSGSPSAEMLKEAMGSDIDLKATLP
AADRSLLPVPMQADQMLDPTVYALVAAGRMIDCEAGCKPMPFTSSVYTVIRPDASF
MLKMSLGVTNSVRNLARLNGDDMRFRHEHLMQPSRSPGLTLIPDAFVGV
KIDGALIDGIVSMRENPFTGANADNVSLLAACEHLDRGSRGLALIRNRAHERR
PLDIYARDWFERFLITFVRPIFGLYLRHGIAMAHQONIVIEIHGYPGLFPAHDER
FPIHERAHGALVEALPGFSPSESVGSEPVDERLLYVAFINSVAGWALGREGVGS
ETALMLRREILRLALRGANGIYRKMALPTLQCKANLKTRLAMDLVGPLETOS
VYIQTNPILPETEKVLAHA"
gene      4673. .5265
           /gene="rhbd"
           /note="previously rhbd"

```

```

CDS
4675. .5265
           /gene="rhbd"
           /function="siderophore biosynthesis"
           /note="similar to IucB"
           /codon_start=1
           /transl_table=11
           /product="Rhbd"
           /protein_id="AAD09415.1"
           /db_xref="GI:4151934"
           /translation="MLRSEPTLAAVSRYPDIPRTISFLLEKQKDLKILMRMNO
AHVQPMKAEIEDIAVYIDNLADPHQDPTIGLIDGTPMSYEAHYAKDDYLCRY
PAEKDRGMHMLVGSPPFGRIAPVIRAFRLFUDPGTQKVAGSPVAARLLR
YAPAKPEEOGIEDLDPDKAKLMFCYRERFIQFGL"
gene      5268. .6632
           /gene="rhbe"
           /note="previously rhbe"
           /note="previously rhbe"
           /function="siderophore biosynthesis"
           /note="similar to Alca"
           /codon_start=1
           /transl_table=11
           /product="Rhbe"
           /protein_id="AAD09416.1"
           /db_xref="GI:4151935"
           /translation="WTDFDLAGIGIGFPNLGLAALLSHENLSNVFLERKPAFRWHG
LILPGTLOVPMADLVMTADPTRLSFLNTLYAVDRLYKPFYFENPMIPROEVHYC
RNASQQLSACRFGEEVVDVAHESASDSITVESRSAGCGKQOYRSNINAIQVTAFFLP
KMAQIKTILAPLMHSSSEGRRLRLSKRRRVVITGGQSAECVLLALDLDTLEMAAG
ASIQWITRSAGFFPMEYSKGLLEYTPPYMRPHRIAPVRREIYADQGLYKGISFS
TIGEIFDLMYERSVCGRDPGLALFENCAVETLESAGSGSFRI GYNHNLDEKATVET
DAIVATGVRHAMPFWMLGSLKGSVLDTCMEGDLVVGDFRARSDDGKHVQNAET
FHHGAGAPDLGIGARNNAVINYQLGRHHYVNASAPQKGLPSSQTA PSSISDFY
AHAS"
gene      6616. .8421
           /gene="rhbf"
           /note="previously rhbf"
           /note="previously rhbf"
           /gene="rhbf"
           /function="siderophore biosynthesis"
           /note="similar to Iucc"
           /codon_start=1
           /transl_table=11
           /product="Rhbf"
           /protein_id="AAD09417.1"
           /db_xref="GI:4151936"
           /translation="MLMHRDPLLPQAFSEELMQVYSKRLAKVIEEFAVERVGVBE
KEGHYRIDIDELOYRFKAKRYVFDNLVDPASLFRHGNSDAPLHDPLAFCAEVLPKL
GVKPMVTAHPIKELENTLVSDAHIAARASKGABLAELDDICMEGETTGH PAVTVSKG
RIGLGYSYLAFTPENRTPTVLTGVSKERASFIABRTLTNEGLVREAVGVSRPESF
CAKLAARGSVYTHMMVYHPQMOMH1VYHRAADIAAGHIVFLCKGDDLYIPQOSVR
TISNLSHEBSSTLKCMTITLNTAVTRGIIGKCALTAAPUTWLDLULARDQPLSEECG
LVLLDERGAMHVVHPQSTIBGAPYQFMENKCMWRDSISALIKSGETGLPAAALLHA
GTDGKRPVQOALKEKSGMTVSEWTAFLPVVIVIPVHLLAKGALASAGQNAVTLILKN
GSRPERLARDPIDIVIVCDOAFPEBASTPEEVRVALICLPAPFLIHFLOTTLFICVR
YMSVLIDORSGLBEHAFWGLARSSILAYOKRPEMAASFATDLDGDEYPRCLNLRVR
LFTHCYADDDERPVDFQGMVDNPLVADKSNAA"
gene      complement(8495. .9439)
           /gene="thra"
           /complement(8495. .9439)
           /gene="thra"
           /function="regulation"
           /codon_start=1
           /transl_table=11
           /product="AraC type regulatory protein Rhra"
           /protein_id="AAD09418.1"
           /db_xref="GI:4151937"
           /translation="METIRPLKFGTLSPDRESRLVCRSILMDLGBATTIAPDEGDLT
GVTGIPMYVYSLATVYVPRMTLVNNSCMGDSGVILRAMDSPLVIRHPIKYBAA
RADVFLPSDASSEITLPEGGRFDCAHLPAVALAKDILKATYKQPLAECPLQQL
TNVAGYLLROEYQSEEHAGMVAHAFYDILLPVLAODIGVSPRETPHNRAASIKMVEQ
NLANGSFSTIDVAERITTPRAIQKFPREGETTFSRYVYIGRSLAKSLILAECATS
ISQIAYNVGFNDLSYFNRTFRSRVGRPSDRLRLAAAA"

```

```

misc_feature 9481..9499
/gene="rhizobactin regulon"
/notes="Pur binding site within divergent overlapping
promoters"
9608..11848
/gene="rhtA"
9608..11848
/gene="rhtA"
/function="Ferrisiderophore uptake"
/codon_start=1
/transl_table=11
/product="TonB-dependent siderophore receptor RhtA"
/protein_id="AAD09419.1"
/db_xref="GI:4151938"
/translation="MGNNNGGISFCVVFVVVIGFTGAVAEQEPANQSEAVTSLEIVV
TGRSAQOISIAITIVVDSQIQAEARSKGTLQOILGETIPSPFASDGARTSFGQ
NLGRPELILVDGVMSARSLSRQDAIDPFNIEVERVLSGATAIYGGNATGGIINI
ITKGDAPFLHAEVTVGMSGFAGSQDPRNAGAVTYNSENWDARLSIAGNRTCA
FYDGSGLIPDITQTSFAFNERIDLMSIGIQIDRRRVEFSQYFDSKQSDYGLY
YGPFFAALADPSLFPETRSYGVEDFPNQRRLSMNVYTDNDVFGQULLQGSYRTERI

Alignment Scores:
Pred. NO.: 2,04e-23 Length: 11853
Score: 453.00 Matches: 109
Percent Similarity: 52.69% Conservative: 67
Best Local Similarity: 32.63% Mismatches: 132
Query Match: 22.13% Indels: 26
DB: 1 Gaps: 9

US-09-889-746-2 (1-400) x AF110737 (1-11853)
Qy 18 AlaValThrGlnGluLeuTyrAlaGlnProAsnGluSerLeuProThrValGluLeuGlu 37
Db 9677 GCTGTGGCAGGAG-----CCAGCGAATCAATCCGAAGCTGTGACGAGCCTGGAA 9727
Qy 38 ProValValIleThrLeuAspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMet 57
Db 9728 GAAATCGTAGTCACCGCGGACGATCC-----GCTCAGCAGATTTCTGAATC 9775
Qy 58 ProHisThrThrLysValIleTyrGluGluGlnIleGlnGluAlaThrGlySerArg 77
Db 9776 GCGCGCACCATTTAGTGTCTGATTCGGATTCAGATCCAGCGCGGCGCGTTCGGGCAAG 9835
Qy 78 GlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGly---ValSerSerGlyThr 96
Db 9836 ACGCTGCACGACATTCGGCGGAAACCATTCGAGCTTCGACCCCGCCAGCAGCGAGCG 9895
Qy 97 ThrSerAsnPheGlyGlnThrMetHisGlyArgGlnValGlnPheLeuLeuAsnGlyVal 116
Db 9896 CGTAGCTCATTCGGCCAGAACCTCGTGGCCGCCGCCACCACTGATCTGTGACGGTGTG 9955
Qy 117 ProLeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAsnProAsnGlnVal 136
Db 9956 TCGATGAACCTCGCGCGCTCGTGTAGCCGACAAATTCGATGCCATCGATTCATCAATC 10015
Qy 137 AlaArgIleGluValLeuSerGlyAlaThrSerIleTyrGlySerGlyAlaThrGlyGly 156
Db 10016 GAACGGTGGAGGTCTGTCTGGGCAACTGCAATCTACGGCGGGAACCAACCGCGCGC 10075
Qy 157 LeuIleAsnIleValThrLysSerAspLeuGluGluGlnPheGluThrArgIleGly 176
Db 10076 ATCATCAACATCATCATCAGGAAGGCAAGGACCGGAAACCG-----GGC 10120
Qy 177 ValHisGlySerLysLeuSerSerGluGlyIleGlyTyrGlnValGlyGln----- 193
Db 10121 CTGCATCGGAAGTAGCCGCGCATGGCGAGCGGCTTTCCGGGAGCAGGACTTCGAC 10180
Qy 194 ---SerValAlaGlyVal-----SerGluAsnGlyAsnValLeuAlaArgLeu 208
Db 10181 CGCAATGCGCGGGTGGTGGTCACTTACATACAGCAAAACATGGGAT-----GCCGCGCTT 10234
Qy 209 AspValAspTyrArgThrThrGlyGlyAlaPheAspAlaAsnGlyLysArgIleAlaPro 228
Db 228

```

```

Db 10235 TCGATCGCGGCAACAGCAGCGCGCGCTTTATGACGGCAGCGCGCTGCTGATCCCC 10294
Qy 229 GluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsnThrAsnValAsp 248
Db 10295 GACATCAGCAGACGCTCACCGCATTCACGAGCGCATCGACCTGATGGGTCTATCGGC 10354
Qy 249 TrpGlnLeuAspAspLysGlnAsnIleAsnLeuAlaLeuThrHisTyrAsnAspGln 268
Db 10355 TACCAGATCAGCAGCAGCGCTGCGTCAATTTTCCGGGCAATATTTTCGACGACGAG 10414
Qy 269 AspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeuPheGlyGluLysPro 288
Db 10415 GACTCGGATTAAGGGCTTTATACGGGCCCTTCTTTGAGCGCTC-----GCCGACCGC 10468
Qy 289 SerLeu---AsnAlaIleLysGlyLeuSerLeuSerGluGlnProLysThrThrLysSer 307
Db 10469 AGCCTGTTTGAACCCGCTTCGGATACGAGTCCGACTTCAATCGCAGACACGCGCTCG 10528
Qy 308 ThrPheAsnIleAsnTyrHisAspAspLeuTyrGlyAsnThrIleAsnThrAsnAla 327
Db 10529 ATGTTGAACGTCACCTATACCGATAATGACGTTTTCGGTCAGCAGCTATTGCTTCAGGGA 10584
Qy 328 TyrTyrArgArgGluLysGlyArgPheTyrProPheValAla 341
Db 10589 TCTTACCGACTGAGCGGATTAAGTTTCATCCCTTCCTGCT 10630
RESULT 19
AP003582
LOCUS Nostoc sp. PCC 7120 DNA, complete genome, section 2/19.
DEFINITION AP003582
ACCESSION AP003582 BA000019
VERSION AP003582.1 GI:17129649
KEYWORDS
SOURCE Nostoc sp. PCC 7120 DNA.
ORGANISM Nostoc sp. PCC 7120
REFERENCE 1
AUTHORS Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
Kaneko, T., Nakamura, Y., Molk, C.P., Kuritz, T., Sasamoto, S.,
Watanabe, A., Iriguchi, M., Ishikawa, A., Kawashima, K., Kimura, T.,
Kishida, Y., Kohara, M., Matsumoto, M., Matsuno, A., Muraki, A.,
Nakazaki, N., Shimpo, S., Sugimoto, M., Takazawa, M., Yamada, M.,
Yasuda, M. and Tabata, S.
Complete genomic sequence of the filamentous nitrogen-fixing
Cyanobacterium Anabaena sp. strain PCC 7120
DNA Res. 8 (5), 205-213 (2001)
2 (bases 1 to 342650)
Kaneko, T.
Direct Submission
Submitted (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research, Yana
1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail: kaneko@kazusa.or.jp,
URL: http://www.kazusa.or.jp/cyanobase/,
Tel: 81-438-52-3935 (ex.2338), Fax: 81-438-52-3934)
FEATURES
Location/Qualifiers
1. 342650
/organism="Nostoc sp. PCC 7120"
/db_xref="taxon:103690"
/note="synonym: Anabaena sp. PCC7120"
complement(224..436)
/gene="as10305"
complement(224..436)
/gene="as10305"
/note="ORF ID: as10305
unknown protein"
/codon_start=1
/transl_table=11
/protein_id="BAB72263.1"
/db_xref="GI:17129650"
/translation="MRAATILAKSHKSFQPPONLAQRSPVAMFRVPVFNDRLLT
PKVTATNRQGLIYFLVFPKYSQSL"
complement(1578..1540)
gene
CDS
gene

```

```
CDS
/gene="a110306"
/complement(578. .1540)
/gene="a110306"
/note="ORF ID:a110306"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="BAB72264.1"
/db_xref="GI:17129651"
/translation="MENISWVGIDVSKATLDVYIRPIGKALKPANTELEIFNLVQL
KPYDLNLIVALEGTGLETLYIQLQAAALPVALLINPROGRNPAKAGKAKTADIDAO
ILAHGBAKPQVYLTIESQASROLGELIRKROLVMQTAERKRNRARAGKADLDEA
HLYEDERLKQNLQIEQLQNNQOMIEKYNLKTTPGIGVYSTLVDSDELGOLT
AKOISRLVAPAPINHDSQGHKEMINGRAHRAVLTMGAVVAMRNHVIKAFYERLL
VERGSKKLALFACVHKMLVIINAMVRDNLPMRVTDNLQIPINA"
/complement(1954. .2298)
/gene="a110307"
/complement(1954. .2298)
/gene="a110307"
/note="ORF ID:a110307"
/unknown_protein"
/codon_start=1
/transl_table=11
/protein_id="BAB72265.1"
/db_xref="GI:17129652"
/translation="MKVLRSIKPEFYISFLYIYQTTMGAGDICIRESVANSGMSK
FIGRQYALAPKGLERDVRANPVIKAVGVGSHKPEHMEMLYEGVDREIALNAL
KPMGFLIRSTD"
2779. .6315
/gene="a1r0308"
/2779. .6315
/gene="a1r0308"
/note="ORF ID:a1r0308"
/codon_start=1
/transl_table=11
/product="5-methyltetrahydrofolate--homocysteine
S-methyltransferase"
/protein_id="BAB72266.1"
/db_xref="GI:17129653"
/translation="MTHPPLKRLHSPELPIYVDGAMGNTLQTNLTAEDFGVQVEG
CNEYLVHTKREAVAKYHRDPLVAGVLEIDTDTGATSYLAIEDLADQTYILAKKAB
LAKVAEFTSTPDKPRFVAGSIGPTTKLPLIGHIDDTLKTCEADQEAALDGVDL
LVEPTCODVLIKALNGIEBEVFKGGERIPLMVSVMEKMTLVGSEBALVDTLEP
FPIDILGNCATGPDLMKPHIKYLAHSPFVVSQIPNAGIPENVGQAHRYTLMPLR
MALMHVEDLGVOVIGGCCGTRPEHIQOALAEIKADLKPVRQBSLEPAASRYTQPY
BODNSPLIVGERINASGSKCCRDILNAEMDGLVSMARSOVKEGHHIIDVNVYERD
GYADHELVSRIVANNVTLPLMDSTWEKMEAGLAVAGCKLINSTNYEDGEPRLAV
LELAKRYGAGVVGITIDEGMARTAKKQIAGRAVQSEYGIPTTELFPDTLALPI
STGIEEDRENGKATIESIRIKELPGCHVILGAVNISFGLNSASPMVANSVLEHAM
TAGMDAIVSASKILPLSKIEEHOEVCROLIDORKFEGDICIYDPLELTLFEGV
TTRKNGVDSESLPIERIKRHIIDGERIGLEQTLKALFOPELAIINTFELDGMKVY
GELFGGQOMLPVLQSAETMKAAYALBPHEKSSGNNAKGVIIANTYKGDYHDIG
KNLVDTILLNSNGYINLGIKQPVENIEAVNKHADCLAMSLGLVKSJAFMEKNLEV
FNEKGINVPVILGGAALTPKFVHKDQNTYKGVIIQKAFSLHFMDKLMPKAKAGV
MWSLSGLFDEVEETEETEPNKSPIPSPOSVPSPQVPIIDTRSEAAVAIDIPRTP
PFWGQLOLQPSDISEEIEFWMHDLQALIAQOMFRKPEKSEKQEAFLNEKVPYLE
TKWRIIAENLHPOVLYGYFPCQSGNTLYVETNSPNAETLTQEPFPROSSKRLC
IADFPAPKSGIITDVPFMOAVTYGEBATERAQCLPANNQTTDYLTFHGILAVQYAKLA
BMTHARIRKELGAEHPDNRIDILQRYQSSYSPGYACPNIQDFQFQDLLETSR
INLYMDSBQLYPEOSTTAIITYHPAKYETA"
6517. .7569
/gene="a1r0309"
/6517. .7569
/gene="a1r0309"
/note="ORF ID:a1r0309"
/hypothetical_protein"
/codon_start=1
/transl_table=11
/protein_id="BAB72267.1"
/db_xref="GI:17129654"
/translation="MQLITTTTFVVLGSLACSISTALFVSSTLAVROLAOSNPSA
VALALAREKNRPIATIVILNNFINIGSLVTGVAIQAQSGQMLGIFSAVLFTLLII
FGBIIRKTLGERYCERISILMAIPITGLSVFPLVWLTLENVAPSKGKRPTNFA
```

```
gene
/complement(7653. .7829)
/gene="a810310"
/complement(7653. .7829)
/gene="a810310"
/note="ORF ID:a810310"
/unknown_protein"
/codon_start=1
/transl_table=11
/protein_id="BAB72268.1"
/db_xref="GI:17129655"
/translation="MKADSDEGGYIGLILDFCNIRGATGLGPGLEQVALGNAPHT
DTENFMRISKTSEP"
/complement(7826. .8197)
/gene="a110311"
/complement(7826. .8197)
/gene="a110311"
/note="ORF ID:a110311"
/unknown_protein"
/codon_start=1
/transl_table=11
/protein_id="BAB72269.1"
/db_xref="GI:17129656"
/translation="MANISCPNCHQOVSQALVCPYCRITLKAQHPGVTLHRAKDT
YICHTCTYHADDTCTPPKRPYAKDCTLYQDLAERNAQLEQCAANTPBAISXWQRN
QNLULLGLLTCFLIVLSTS"
8416. .10179
/gene="mend"
/8416. .10179
/gene="mend"
/note="ORF ID:a1r0312"
/mend_gene_product"
/codon_start=1
/transl_table=11
/product="2-succinyl--6-hydroxy-2,
4-cyclohexadiene-1-carboxylate synthase"
/protein_id="BAB72270.1"
/db_xref="GI:17129657"
/translation="NSNLMPIAYKINQNLMAVITFETLKRGLAVAYICPGSRSTPLA
VAFQAQAPRTEIASILDSBSAEPFALGIKATNRVAVCTSGTAGNFPYAVIQAOS
EVSQVALLTADPELRLDCHSGQTTDKMLYGSYNNQOAEALPVSQMGMLAYLQTL
VSWTRMOAPRPGFHLNIPRPDPLAPIPDGTDLGYLAKPHBEPFAGITPTTPLPH
HSPRLTIPHEMLOSQGIITIAAGAOQOQYQYERARARLSQTLQWVYLABGSLPVNRY
DNPVLTISTYDILINQOATRLAPDVYIQIDMPTSKELRWITDHQRRRVIDPSD
QNLDELRGATLHRLRVEBELGYQVEBKSVSSEYIQLMCNAETVRVAVVDDTLKME
DLVECKAAMLQMLPPEPLPIANSMFVRDVEFPWKNNLRLVRSHFRNGANGIDGTL
STALGISHRQSSVYITGDIALHDTNGFLIRNKFEVGHLLIILNNGGIGPEMLPIA
KEBPBEPBEPPOIDIPAOICTTYNVQHELHSHVHLQOQRLNPLPNTGIRVLELRIN
RKIDQWRKDNISNFAADNIT"
/complement(10282. .10614)
/gene="a110313"
/complement(10282. .10614)
/gene="a110313"
/note="ORF ID:a110313"
/unknown_protein"
/codon_start=1
/transl_table=11
/protein_id="BAB72271.1"
/db_xref="GI:17129658"
/translation="MEMPOQSVRYKQDSYNINELRLKTFCLGMKIMEBOTKPKTEL
YEFITLACLOVDQRQDFITSLFIELADMIKRTSLNRNKKEDLITAPRREFPALADP
EDIAVXSN"
/complement(10805. .11107)
/gene="a110314"
/complement(10805. .11107)
/gene="a110314"
/note="ORF ID:a110314"
/codon_start=1
/transl_table=11
/product="transposase"
```



```

gene
/translation="EENPCQQLLTGNGNGWKKKFGDR"
/length=1577.2098
/genes="s1r0613"
/length=1577.2098
/genes="s1r0613"
/notes="ORF_ID:s1r0613
unknown protein"
/codon_start=1
/translation_table=11
/translation="MSYLLAVANRIAEAVTTLEQAGFAOKNLTITIGTKYVADF
GLVDPKKQAIKRAKLMATLVPFGPAAGCFNLITGLSTLDMAQDPGNHIVGGLGAI
GGTMSFPFVGAGVGLSFGSGDSLPRYNLLQAGKYLAVVAGSELQKQRAITNLRPLNPE
YLOGYAPDEBAFV"
complement(2172.2873)
/genes="s110558"
complement(2172.2873)
/genes="s110558"
/notes="ORF_ID:s110558
hypothetical protein"
/codon_start=1
/translation_table=11
/protein_id="BAA16582.1"
/translation="MSDNLTELSQQLHDSSEKKQLTAIALAMEGEGGQIILLDYLA
K NPLKRPVLAVGNVQTLNLRBOETTLOLRNYPIQSAOGIDVLPLOEALG
ODPRTADEITRDLCELAGRGASQRMVLTFTVEKRPALDHTINATLWHLHSGNNGP
SVQRRLMASGSEFTKLPKPIGMKSGNVTWPKGFTWDLASQGHLPILNIGVRA
ASLYRHPVMSQYGM"
complement(3192.4268)
/genes="s111214"
complement(3192.4268)
/genes="s111214"
/notes="ORF_ID:s111214
pN134 or A1103 phytochrome-regulated gene homologue"
/codon_start=1
/translation_table=11
/protein_id="BAA16583.1"
/db_xref="GI:1651655"
/translation="WYNTLEKPGPDEIRGCVTKPAKETTLTPFYTTDDENAKNDIS
PNDEIRALIEBRVDYNNRHHFVNRNPFNSFNKSWDHIDGKRLQLEVEFLERSCTAFSGF
LYXELGRRLKNGRPLLAECFNLMSRDEARHAGFLKAMSDPMLSLDGLTKSRXT
FEKPKTIFVATYSEKIGYWRVITIRHLKPNDCIYPIFEFENNCOENHGPFP
DAIRAPQHTLNDWKAKMCRFPLSVFATMYNDORADFYACIGLEARSYKXVTE
KIMETAGRVFPTILDVNNPEFTNRLETCVSNNEQLAIDASGARVYIKALRKLPITAS
NGWQTKLYMKRIAVDQLAGVR"
complement(4451.5389)
/genes="s111213"
complement(4451.5389)
/genes="s111213"
/notes="ORF_ID:s111213
hypothetical protein"
/codon_start=1
/translation_table=11
/protein_id="BAA16584.1"
/db_xref="GI:1651656"
/translation="WLSLEKORILVTGAGCFLKQVVAQILAAQGRPKTITPRKDC
DILQASACERAVENODIVHLLAAVGGIIGLRKRPALPYDNLIMVQVLIHAHQAGV
KKEVCVGTICAYPKPTVPVPEKEDLMNGVPEETNAPYGAKKALLVQLESYRLQYGFN
GIYLLVNLGPEPDNPPRSSHVI PALIHVYEAQAGQQLPVMGDGSPTRFELST
DAIRGIVMGTOAVDADVPVLTGTFEISIKDLTELCELMPEFGDIIMETDQNGQPR
RCLDJTKAKERGFPAOVSLRGLKNTIDMYRQRS"
complement(5534.6622)
/genes="r1bD"
complement(5534.6622)
/genes="r1bD"
/notes="ORF_ID:s111212"
/codon_start=1
/translation_table=11
/product="GDP-D-mannose dehydratase"
/protein_id="BAA16585.1"

```

```

gene
/translation="MSKSKVVLNGITGQDSYLSLELLKQYVHGIIIRITSTPNTD
RIDLHYVPHDLKALRIHAGDITDGTTLRLIEDEKTEIYINLGAQSHVRSFSPSE
YTVDSVAMGTILRLBAIRDYORRTGIQVRFQAGSGSEMGKVQBIPOKETTPFVPSRP
YACAKVGMOTVNRRESYDLFACNGIIPNHSPPRGSTPRKTRITRAIRVACTOK
KLITGNTSKRDMGTAKDYVRAMAMLOQOEBDDYVVAITGETHEKEPELIRFGVNL
NMQNVAFDERLYRPAEVDLLIGDPARKQAQGWBSVTFELVHLMVADLAVGLT
SPNQSGRIRKELMAQMAPRIRSONGHAVD"
/length=7229.8311
/genes="psba2"
/length=7229.8311
/genes="psba2"
/notes="ORF_ID:s1r1311"
/codon_start=1
/translation_table=11
/product="photosystem II D1 protein"
/protein_id="BAA15586.1"
/db_xref="GI:1651658"
/translation="WTTTLQQRBSASLWEQFCQVTSITNNRIYGVGFGLMIPITLLTA
TTCFIATAPPVVDIGIRBEPVAGSLYGNNIISGAVVPSNAIGLHPYPIWEASL
DEMLYNGGPYQLVPHPLIGIFCYGRQWELSYRGMPEMPCVAVSAPVATVFLI
YPIGGSPSDGMPDIGSTPNPMIVFOAEHNILAMPHMLGVAGVFGSLFAMHGLI
VTSSLVRETTVESQNYGKRGQEBETNYIAAHGYPGRLLIYQVASFNNRSIRHFFLG
AMPVIGIFPFWAGVSTMAFNLGNFNGSITLDSQGRVIGTMDVILNRANIGFVWHER
NAHNPFLDASGEQAPVALTPAVANG"
/length=8492.10471
/genes="spea"
/length=8492.10471
/genes="spea"
/notes="ORF_ID:s1r1312"
/codon_start=1
/translation_table=11
/product="arginine decarboxylase"
/protein_id="BAA15587.1"
/db_xref="GI:1651659"
/translation="WGEPPVPADKALGKKPKKKNASWSIBESALRYEAMCAPYFAI
NAAQGITVSPNGDRGSDLELVEALQRKGLPLIRFSDILDERLWSCPKA
IARVYVPTYQAVYVVKCQOGRHLEVALVRFGQTCQGLGASKPELMIALATPLPL
DRQDKTKPLITCNGYKDDYLETALMLAKRHP111IEQLRELMVLIISQNLK
PMLGVRARLSQSLKRSISGNGDRALKGLPDIPIVYIRLEENNCIDCLKMHF
LCTQVSDIALIREANRBSQIYVEIKGAKRRYIANGGGLAVDYGSTNYPASKNY
NMQANDIVAIQACEIGQVSPITLVSESBRATMAHQSVLPVIVLSNQTGSEPH
PDBENAPHLKMLWECYETITAEQYQEOYHDLQIKTESLSLFGYSLTTRGABQ
IMHACCRKIFETRLQLEYIPEDFOALDKIMTDIYYVNLVSFQSAEBSLDDLPFLP
IHLHLEKSORVITLADLDCDSQDKIDRFIDLVRSYLEVHLENDGNPYIGMFLVG
AYQELMSNLHNFQDINNVHIAITTPQGOVIBSVVAGDPTMTVLGVGVQSDDLLEGLR
RHTIELNSGQITLESRRLEDEYQSIRRYTYS"
complement(10622.12631)
/genes="11g"
complement(10622.12631)
/genes="11g"
/notes="ORF_ID:s111209
gene:11g or 11gA"
/codon_start=1
/translation_table=11
/product="DNA ligase"
/protein_id="BAA15588.1"
/db_xref="GI:1651660"
/translation="WTTPEKRLIQRLQOLKASYAYVLDAPVMEQSVYDQLYREIQR
EAKNPELITPDSPTQVGEQSPASQFRSVAHNPIVLSLENAFVQLOQOERWQRIAP
TIEKAEVCELIDISALATYENGLVAGVTRGGTGEESQNIKITRSIPVKNL
DNPPTVVEKGAFLPLEEFNRINNEREKGSSLPANPNAAAGTLROLDPKIVHOR
LOFPAYTTLHPQEDKIQSQOMALCYELKKGAGSNPHQOLCGDLGEVVAIPFDMGAR
QALPYMTDGVVYKINQYPLQREBELGTQKPRRAIALKPAEETPTVYVVAIEYVNRG
ATVPLAVMEPVOLAGTIVORATLHNOIRIQSIDIRVGDVLIIRKGEI11PVRVWTE
IAPENTPYIIPSHCPAGCSPLVRPLEEAVIRCVNMSGALIGSLIHMASNALDIQ
GLGEKRVITLLENRLVNSVADLYGQVEQLGLERFPAQGSARKLI1A1EVSQSPMR
IIFGIGIRHVQGVNAKLSQCPYVEKISOASIPLEGVYIGSPRIAEAVVWMPRPG
NOOILQDLBEIGLIVIANOGIDOTKTDGSLKGTFTVLTLPNLSRLAEQDELIEQSG
KYTSSVSTTIDVYLLGDRGSAARAEISLIGITLSBAERTQLLEF"
/length=12753.13363

```

Alignment Scores:

```
Pred. No.: 9.24e-22 Length: 133859
Score: 448.50 Matches: 113
Percent Similarity: 52.63% Conservative: 67
Best Local Similarity: 33.04% Mismatches: 143
Query Match: 21.91% Indels: 19
DB: 1 Gaps: 10

US-09-889-746-2 (1-400) x D90899 (1-133859)

Qy 22 GlnLeuTyrAlaGlnProAsnGlnSerLeuProThrValGluLeuGlu----- 37
Db 19952 CAATTGCTGATTCGACCGAGGGTTCCGGGCAACCGCGAGATAGAGAAAGTACAGGCC 19893

Qy 38 ---ProValValIleThrIleAspLysSerGlyMetAlaLeuAlaAsnArgIleThrGln 56
Db 19892 CTACAAATTTGTTCCGCCACCGCTACAGAGAGGAAATCGCTAAC----- 19845

Qy 57 MetProHisThrThrLysValIleTyrGluGluGlnIleGlnGluGlnAlaThrGlySer 76
Db 19844 ATTCCCGGTCGGTAACGGTAATTGACCGGGCAGAAATTTGACCAACAACCCAGGTTTAC 19785

Qy 77 ArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSerSerGlyThr 96
Db 19784 TCTTCCTGGCGGATTTTGGGTCAACTAGTGCAGGCTTAGCCCCCAGCACCGGGAGT 19725

Qy 97 ThrSerAsnPheGlyGlnThrMetHisGlyArgGlnValGlnPheLeuLeuAsnGlyVal 116
Db 19724 GCCAGTCAATTTGGCCAGGCTTTGCGAGGACGTAACTTTGTTGTTGATGACCGGTG 19665

Qy 117 ProLeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAsnProAsnGlnVal 136
Db 19664 CCCCCAACCAACCGTAATGCTTCCAGATTTTGCAAACCAATGCTCCCAAGTCCCAT 19605

Qy 137 AlaArgIleGluValLeuSerGlyAlaThrSerIleTyrGlySerGlyValAlaThrGlyGly 156
Db 19604 GAACGATTTGAAGTATCAGGGGCGCAACGGCCATTTACGGGATGGTGCCACTGGAGA 19545

Qy 157 LeuIleAsnIleValThrLysSerAspLeuGluGluGlnPhe---GluThrArgIle 175
Db 19544 GTAATCAACATCATCACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 19485

Qy 176 GlyValHisGlySerLysLeuSer---SerGluGlyIleGlyTyrGlnValGlyGlnSer 194
Db 19484 GCCATCAACACCGCATTTTGCAGAGTGATATCAACAGCGCTAGGGAGGATGCTGGAACAAT 19425

Qy 195 ValAlaGlyValSerGluAsnGlyAsnValLeuAlaArgLeuAspValAspTyrArgThr 214
Db 19424 GTGGCGGGAACCTTGACTATGCTGACTAC-----GCTTTCACCGCATCCTATGAAATCC 19371

Qy 215 ThrGlyGlyAlaPheAspAlaAsnGlyLysArgIleAlaProGluPro---AlaGlnThr 233
Db 19370 GTTGGGGGATTTTGCACGCTTAGTAATCGATTCGATTCCTCCCGCATCCCAATGGCCAAG 19311

Qy 234 AspLysGlnAspSerLysSerLeuSerValAsnThrAsnValAspTrpGlnLeuAspAsp 253
Db 19310 CGAGTATCGACACTGATAGCTTAATGTTTGGTGAATTTGGGCAATCGGCATCAACATGCTAC 19251

Qy 254 LysGlnAsnIleAsnLeuAlaLeuThrHisTyrAsnAspLysGlnAspThrAspTyrAla 273
Db 19250 GAACAGCGCTCGAAATACCAATTAATCAATTTTCAAGCCCAACCAACCACTGATTCACC 19191

Qy 274 ProAspTyrGlyAsnArgLeuAlaValLeuPheGlyGluLysProSerLeuAsnAlaIle 293
Db 19190 GTTGAT-----CCTAGTATTACGGCGATCGCGGTAGACAGAGATCC---CAAGCGATT 19140

Qy 294 LysGlyLeuSerLeuSerGluGlnProLysThrThrLysSerThr---PheAsnIleAsn 312
Db 19139 GACGGCTGGACCTT---GATACCCCGCAACCAACCAATTAACACCGTAGCTAGCCTTGAT 19083

Qy 313 TyrHisAspAspLeuTrpGlyAsnThrIleAsnThrAsnAlaTyrTyrArgArgGlu 332
Db 19082 TACAGCCACAGTAACGCTCTCAATGTCACACCTTAGGGGCAAAATCTACTATCTGATTTAT 19023
```

```
Qy 333 LysGlyArgPheTyrProPheValAlaProPheSerIleAlaLysAlaLeuProIleLeu 352
Db 19022 CTGACTAGGTTTTTCCCTTCGACGCGGAGCCTCCGTTAGCTTGGGCAATAGTATTTT 18963

Qy 353 GlnSer 354
Db 18962 CAGTCT 18957

RESULT 21
ABO10890 2734 bp DNA linear BCT 20-FEB-1999
LOCUS Vibrio orientalis gene for iutA, complete cds.
DEFINITION ABO10890
ACCESSION ABO10890
VERSION ABO10890.1 GI:4239822
KEYWORDS iutA.
SOURCE Vibrio orientalis (isolate:SD004) DNA.
ORGANISM Vibrio orientalis
Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE 1 (sites)
AUTHORS Murakami,K., Fuse,H., Takimura,O., Inoue,H. and Yamaoka,Y.
TITLE Cloning and characterization of iutA gene from Vibrio species
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2734)
AUTHORS Murakami,K.
TITLE Direct Submission
JOURNAL Submitted (03-FEB-1998) Katsuji Murakami, Chugoku National
Industrial Research Institute, Marine Biological Technology
Section; 2-2-2 Hiro-suehiro, Kure, Hiroshima 737-0197, Japan
(E-mail:murakami@cniri.go.jp, Tel:81-823-72-1938,
Fax:81-823-73-3284).
FEATURES
source
1.2734
/organism="Vibrio orientalis"
/isolate="SD004"
/db_xref="taxon:28175"
330..2447
/genes="iutA"
330..2447
/genes="iutA"
/codon_start=1
/transl_table=11
/product="iutA"
/protein_id="BA074703.1"
/db_xref="GI:4239823"
/translations="MTVKGSFPLSTIALGVAAATTVFAAQAEYTSSEKVVVSSHM
PKASIDIPGTWVIDSETIQEYRGKSLDEILLAATIPSLDVSNGRTHSGQNLGRS
LMWIDGVSLSQSVSLSRQSDSDPENNDRIVLSGATSIYGAGSAGVINIITKKAS
GEVERESFVSGTSGVSSSEDFYKLAQSTAGGNEKVGASVAYTKQTFPDADGEM
VTDPITGQSTQFNETIDLLGNQINIAEGQNLNLAQYDSQQDTPYGLIFERKDSGN
FQFVDRREGYSADREOQTERVMSAAYSNNFNLSHQLIAELSYRTEDTFMPYVDGK
STLSTNSRSQSTDI LAMKAALNKSFNANLTYGIDYIDKPSDKAIYDQTTTATGG
LVHVISMOGRYPGIDTQS IAGYIOAEYNITDDWSVOTGYRQKMDIDVDFQATSSS
DETPGSDTYAEGFLNGLTYHINPTSOVWANSQGEDIPOLYRYGKSKATODVAASK
LEGITDSYIEIGYRNLDNLNLSQASIIYSSHSATVYEDSLISGLADPKRYGAEAQ
ISWATDMLQGLUHYVVTETDITDGDHDKAMEASTSKAGAWGWDNDYSLKLOS
LTWFDYEDADQRELDGFTVDFVTGYNQVLPVGLPGSLGFKINLFNEDYTTTWSQAHWT
GSDALYQYKGRGRTFTLNYQKF"
```

```
BASE COUNT 797 a 636 c 618 g 683 t
ORIGIN
Alignment Scores:
Pred. No.: 2.47e-23 Length: 2734
Score: 441.00 Matches: 115
Percent Similarity: 53.37% Conservative: 67
Best Local Similarity: 33.72% Mismatches: 129
Query Match: 21.54% Indels: 30
DB: 1 Gaps: 8
US-09-889-746-2 (1-400) x ABO10890 (1-2734)
Qy 11 SerLeuProLeuLeuSerValAlaVal-----ThrGlnGlnLeuTyrAla 25
Db 348 AGTTTCCCACTCTCCACTTCGCACTTCGGCTCGCTGCTGCGACCAACCGGTATTTCCCGCT 407
```

QY	26	GIProAnsgluSerLeuProThrValGluLeuGluPProValIValIleThrIleAspLys	45
Db	408	CAGACACAGAAATAC-----ACCAAGCGAAGAAAAAGTGTGTGTTCA-----	452
QY	46	SerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrThrIleValIleTyr	65
Db	453	AGCCCAATGCCCTAAAGCG-----ATCAGCGATATCTCCGGAAACGTTGGTATATCAT	506
QY	66	GluGluGlnIleGlnGluGlnIleThrGlySerArgGlnLeuAlaAspValMetAlaGln	85
Db	507	TCAGAAACTATTGACCAAGAAATACCGCGGTGTGTAAAGCTTGATGAATATTTGGCTCA	566
QY	86	LeuIleProSerLeuGlyValSerSerGlyThrThrThrSerAsnPhenGlyGlnThrMetHis	105
Db	567	ACCAATCCCTTCACTGGATGTAAAGACGCGCGGTGCTTACCACTGCGCCAAAACCTACTGT	626
QY	106	GlyArgGlnValGlnPheLeuLeuAsnGlyValProLeuThrGlySerArgAspIleSer	125
Db	627	GCTCCACACCTTAATGATATGATATGATGATGATATGATATGATATGATATGATATGAT	686
QY	126	ArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyVal	145
Db	667	CGTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	746
QY	146	ThrSerIleTyrGlySerGlyValAlaThrGlyGlyLeuIleAsnIleValThr---LysSer	164
Db	747	ACTTGATCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	806
QY	165	AspLeuGluGluGluGlnPheGluThrArgIleGlyValHisGlySerIleLeuSerSer	184
Db	807	TCTGCGCAAGAAAGTAAATTTGAATCATTTGTTAGTGAAACGTCCGGGTCTCATCCGAC	866
QY	185	GluGlyIleGlyTyrGlyValGlyGlnSerValAlaGlyValSerGluAsnGlyVal	204
Db	867	GAAAGTTTGTACTACAAACTGCGCTCATGTTATGCTGCTGCTGCTGCTGCTGCTGCTG	920
QY	205	LeuAlaArgLeuAspValAspTyrArgThrThrGlyValAlaPheAspAlaAsnGlyLys	224
Db	921	CAAGGTCCGGCATCTGTGCTGCTTACACAAACAAACAAAGCTATTGATGCTGATGCTG	980
QY	225	ArgIleAlaProGluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsn	244
Db	981	ATGTCAGACGCCGATATTAATCTACGAGGCTTCAACCCAGTTTAAGAAACCATGATCTACT	1040
QY	245	ThrAsnValAspTyrGlnLeuAspAspLysGlnAsnIleAsnLeuAlaLeuThrHisTyr	264
Db	1041	GGTAACCTTCAAAATTAATATATGCTGAGGTCAAGCATTAATTAACCTGTTAGCGCAATCT	1100
QY	265	AsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeuPhe	284
Db	1101	GATACCAACAAAGACACACCATACGCT-----CTGTAC	1133
QY	285	GlyGlyLysProSer-----LeuAsnAlaIleLysGlyLeuSerLeu	298
Db	1134	TTTGAAGAAAGATGAGCAGCGCTTAACCTTCCAGTTTGTCGATGATGAGAGGCTTACAGCA	1193
QY	299	SerGlnGlnProLysThrThrLysSerThrPheAsnIleAsnTyrHisHisAspAspLeu	318
Db	1194	GATCGGACCAAGCAGCACAACGTCATGTTGAGGCGCCATACAGCAACATTAATCTTC	1253
QY	319	TrpGlyAsnThrIleAsnThrAsnAlaTyrTyrArgArgGlyLysGlyValArgPheTyrPro	338
Db	1254	TTAAGCACAACCACTATGCTGAGTATCTTACCGATAGTAAGATCATACCTTTATGCGG	1313
QY	339	Phe 339	
Db	1314	TAT 1316	
RESULT 22			
AB074151		3495 bp	DNA linear
AB074151			BCT 08-MAY-2002
DEFINITION			Vibrio vulnificus genes for putative transcription regulator.

```

ACCESSION      ferric aerobactin receptor homolog, putative aldolase, partial and
                complete cds.
VERSION         AB074151
KEYWORDS       AB074151.1 GI:20502465
SOURCE         Vibrio vulnificus (strain:M2799) DNA, sub_clone:pV101.
ORGANISM       Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
REFERENCE      1
AUTHORS        Aso, H. and Yamamoto, S.
TITLE          Cloning and characterization of the Vibrio vulnificus gene encoding
                the ferric aerobactin receptor
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 3495)
AUTHORS        Yamamoto, S. and Aso, H.
TITLE          Direct Submission
SUBMITTED      10-NOV-2001 Shigeo Yamamoto, Okayama University,
                Faculty of Pharmaceutical Sciences, 1-1-1 Tenshima-naka, Okayama,
                Okayama 700-8530, Japan
                (E-mail:syamamoto@pheasant.pharm.okayama-u.ac.jp,
                Tel:81-86-251-8473, Fax:81-86-251-7926)
FEATURES       Location/Qualifiers
                1..3495
                /organism="Vibrio vulnificus"
                /strain="M2799"
                /db_xref="taxon:672"
                /sub_clone="pV101"
                .c1..459
                /codon_start=1
                /transl_table=1
                /product="putative transcription regulator"
                /protein_id="BAB91402.1"
                /db_xref="GI:20502466"
                /translation="RLIFEPAGLLEPSWMLDINALRQCRRDRLHQAIOQPDPTVASDLF
                LSGDPEHILACSGNRFILGNMQHNRMKATDILVSMHLOTSVIKSCGRLEILL
                LKGNKRAAKKLQTLLENDIRVWQRTSVNTHLSLSEKLVNSITQDTI"
                496..525
                715..720
                735..752
                /bound_molety="Fur protein"
                738..743
                824..827
                834..2981
                /gene="iutA"
                /gene="iutA"
                834..2981
                /gene="iutA"
                /note="ferric aerobactin receptor (iutA) homolog"
                /codon_start=1
                /transl_table=1
                /product="ferric aerobactin receptor homolog"
                /protein_id="BAB91403.1"
                /db_xref="GI:20502467"
                /translation="MTASKSKPPLSTWALAIIASTALTAQAQSTINDKMAVVVSRT
                PKASDIPGTWYIESDKIRIQEYRGCTTGDIILASATPSIDSGARTNGVLRGK
                MLWVIDSVLOSRSIRLHSDIDFENVDIRIVLVGSTSYTGASGSGVINITTKAH
                SEELSEGVGSGTSGFNSSDFPKVAQSIGNEKVAQTSVYETGGFPFANGTI
                VTPIIDSGQLQFNQTDVFLTVAGDVSETKLVNSSAIQDPSYGLIYGRDFVARK
                GFSEDRSGEGRKIMLSASVADAPLGGQILAEASRKEDQTPPYQSSQQLTDVVS
                LKALAKSPDKINIVGIDVAYODKLNSQALVYOPTIANNNGNLVNTKYAOVGRPGVE
                VSVTAAPQVDVYAITDMPRVEGGRVQYMLNKIDDPGVNQOKKIALDSTSDAVG
                GETDYVALFRNLGTYHLNRDSQVMNAPSGCFDLADPARKYGGGRTLVLDHNDLDS
                INNVQSGIKTNSVELGRIVASGVNLTQTAIVSSQDNIRKNTTLLIELEEDKK
                RYVQLBAMLSYVWVDNIQLGASGHYVSVYKCGKMGEDYDAGCAVASTSKAAMWGSDA
                NNVKLOSMTFVEDASNNKLNQYTLFDLVGSGIPVSGIGFQIQLVLDKDYTTAWG
                ORALILASAYESAAYVYKRGRTYTLNVQVK"
                3001..3021
                3176..3193
                3196..3225
                complement(3241..32495)
                /codon_start=1
                /transl_table=1
                /product="putative aldolase"
                /protein_id="BAB91404.1"

```

```
/db_xref="GI:20502468"
/translation="VDIEVLHGASGVDEWVRRCIELGVCKVNVVATELKIAFADAVK
GHFAHPDANPRKTYITPGKAMKRVKMDKIMCGSEGLK"
BASE COUNT 942 a 923 c 787 g 843 t
ORIGIN

Alignment Scores:
Pred. No.: 1,558-22 Length: 3495
Score: 432.00 Matches: 130
Percent Similarity: 49.76% Conservative: 78
Best Local Similarity: 31.10% Mismatches: 161
Query Match: 21.10% Indels: 50
DB: 1 Gaps: 11

US-09-889-746-2 (1-400) x AB074151 (1-3495)

Qy 7 PheGlnTrpLeuSerLeuProLeuSerValAlaValThrGlnGlnLeuTyrAlaGln 26
Db 855 TTCCTCTATCAACCATGCGACTTGGATTATCGCGCAAGTACTGCACTGATGCCCA 914
Qy 27 ProAsnGluSerLeuProThrValGluLeuGluProValValIleThrIleAspLysSer 46
Db 915 GCACAAAGAA-----ACCTCAACCAACGACCAAAATGGTGGTTTCAAGCGCGACT 965
Qy 47 GlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrThrIleValIleTyrGlu 66
Db 966 CCAAAAGCG-----ATCAGCGATATTCCTGGTACGGTTTGGTATATCAGATCC 1013
Qy 67 GluGlnIleGlnGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeu 86
Db 1014 GACAAATTTGACAGAAATACCGCGGTGTAAGACTCTTGGCGATATTCGCACTCGGCA 1073
Qy 87 IleProSerLeuGlyValSerGlyThrThrSerAsnPheGlyGlnThrMethIleGly 106
Db 1074 ATTCCGCTCTCGAGCTCAGCAGTGGCGTCGCACCACTAGGCCAAACCTCCGTGGG 1133
Qy 107 ArgGlnValGlnPheLeuLeuAsnGlyValProLeuThrGlySerArgAspIleSerArg 126
Db 1134 CGTAAGATCTCGTAAATGATAGACGCGCTTCTTCGCAATCTTCACGCTCCATCAGCCGT 1193
Qy 127 GlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyAlaThr 146
Db 1194 CATCTTGATTCATCGATCCATCAACGTTGATCGTATCGAAGTGTGTCGTGCTCCACT 1253
Qy 147 SerIleTyrGlySerGlyAlaThrGlyLeuIleAsnIleValThr---LysSerAsp 165
Db 1254 TCCATCTATGGCGAGCGCATCGGTGGTGTGCATCAACATCATCAACCAAAAGCGCAC 1313
Qy 166 LeuGluGluGluGlnPheGluThrArgIleGlyValHisGlySerIleLeuSerSerGlu 185
Db 1314 AGCGAAGAGCTTGAGTTGAATCGTATGGCGGGGACTTCTGGCTTTAACTCCAGCGAA 1373
Qy 186 GlyIleGlyTyrGlnValGlnSerValAlaGlyValSerGluAsnGlyAsnValLeu 205
Db 1374 GATTTTGACTACAAAGTGGCTCAATCTATTCTGGCGCAATGAA-----AAAGTTCAA 1427
Qy 206 AlaArgLeuAspValAspTyrArgThrThrGlyValAlaPheAspAlaAsnGlyLysArg 225
Db 1428 GCACGCACCTCTGTGTTTACACCGCAACACAAAGGTTCTTTGATGCCAAGCGCACCAT 1487
Qy 226 IleAlaProGluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsnThr 245
Db 1488 GTGACACCGGATATCTCAACAGGCTCAGTGCAGTTTAACACAGCGCTCGATTTCTTAACC 1547
Qy 246 AsnValAspTrpGlnLeuAspAspLysGlnAsnIleAsnLeuAlaLeuThrHisTyrAsn 265
Db 1548 ACGGTGGCGTTGACGTGTCAGACGAGCAAAAGCTCAAC-----TATTCCT 1592
Qy 266 Asp-----LysGlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeu 283
Db 1593 AGCGCAATACAGCAAGATTCGCGCATAT-----GGCCTGTAC 1628
Qy 284 PheGlyGluTyrProSerLeuAsnAlaIleLysGlyLeuSerLeuSerGluGlnProLys 303
```

```
Db 1629 ATTGTGGCAGAGACTTTGTTGATGTGCGTAAGGCTTTGAACTCTGACCGTAGCAGCGC 1688
Qy 304 ThrThrLysSerThrPheAsnIleAsnTyrHisAspAspLeuTrpGlyAsnThrIle 323
Db 1689 ACCAAGCGCATATGCTAAGCGCCTCTTATGCTGATGCACAGTTCTTGTGGCAGCAGCTC 1748
Qy 324 AsnThrAsnAlaTyrTyrArgGluLysGlyArgPheTyrProPheValAlaProPhe 343
Db 1749 ATCGCAGAAAGCTCGTACCGCAAGAGACCAAACTTACACACCTTATATCAA----- 1802
Qy 344 Ser-IleAlaLysAlaLeuPro----- 350
Db 1803 TCTTCTGGCCAAACAGATTACCGACGTGATCTCGTCAAGCGGCAGCTGCGCAAAAGCTTC 1862
Qy 351 -----IleLeuGlnSerMetAsnLeuProSerAlaThr-----LeuAs 363
Db 1863 GACAAGCTCAACATTTCTATGGCATCGATCGCTACCAGACAGCTCGACAGCATCAA 1922
Qy 363 pAlaTyrThrLysAlaProGlnAlaArgAlaTyrGlyValLeuGlnSerGlySerLysAl 383
Db 1923 GCGCTTTACGATCCACCATC-GCTAACAAATTCGGCAATTTGGTGAACAAGACTTACGC 1981
Qy 383 aGluValLeuGlyArgValProAsnLeuAsnLysProLysArgAlaLeuPhe 400
Db 1982 CCAAGTG---GGCGTTATCTCTGCGTGAAGTAGTGCTGCGTGGCGCTTTT 2030

RESULT 23
AP003589 341880 bp DNA linear BCT 28-NOV-2001
LOCUS Nostoc sp. PCC 7120 DNA, complete genome, section 9/19.
DEFINITION AP003589 BA000019
ACCESSION AP003589.1 GI:17131372
VERSION 1
KEYWORDS
SOURCE Nostoc sp. PCC 7120 DNA.
ORGANISM Nostoc sp. PCC 7120
REFERENCE 1
AUTHORS Kaneko,T., Nakamabe,A., Murakami,Y., Wolk,C.P., Kuritz,T., Sasamoto,S.,
Kishida,Y., Kohara,M., Matsumoto,M., Matsumoto,A., Muraki,A.,
Nakazaki,N., Shimpo,S., Sugimoto,M., Takazawa,M., Yamada,M.,
Yasuda,M. and Tabata,S.
Complete genomic sequence of the filamentous nitrogen-fixing
Cyanobacterium Anabaena sp. strain PCC 7120
DNA Res. 8 (5), 205-213 (2001)
21595285
2 (bases 1 to 341880)
Kaneko,T.
Direct Submission
Submitted (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research
Institute, The First Laboratory for Plant Gene Research, Yana
1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:kaneko@kazusa.or.jp)
URL:http://www.kazusa.or.jp/cyanobase/,
Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)
Location/Qualifiers
1. 341880
/organism="Nostoc sp. PCC 7120"
/db_xref="taxon:103890"
/notes="synonym:Anabaena sp. PCC7120"
/complement(60..758)
/gene="all2281"
/complement(60..758)
/gene="all2281"
/notes="ORF ID:all2281"
/codon_start=1
/transl_table=11
/product="two-component response regulator"
/protein_id="BAB73980.1"
/db_xref="GI:17131373"
/translation="MTEYILPVDDEPLIKELVTHPRHKIQCEYEYIFAHNGDLE
KLANQFIDLLVTDINMPEIDGLTLLSKIRDLDIKAVISAYEDLKNIRRNQQA
```



```

gene
CDS
    complement(751..6102)
    /gene="a112282"
    /note="ORF ID:a112282"
    /codon_start=1
    /transl_table=11
    /product="serine/threonine kinase with two-component
    sensor domain"
    /protein_id="BAB73981.1"
    /db_xref="GI:17131374"
    /translation="MWKLTGYOIVEKIYDSANSEYRSIROENDOSVILIKVLOEYPT
    PIELFRKOYEETIRNLDFDGYIKAVSLSEPYRTVYLILEDFGDTIDRPHNINISIEK
    FLIAIKTPTDIAQIHAANVHKDINPSNIVNPKTGKIKITDGPATPTRENTTK
    HNVLEGTLAYISPEQIGRMNRALDRIDTYSIGVTFYELTKQLEPDTTSLLEVYC
    HIAKQPIPEHEIHSNIPKTLSDIYMKLAKTAERYQSAGIKADLEFCLTGYQSGKN
    TSFSLGSDISDKFQIPQKLYSGEGEINNLTAFTVSGQTEMMLVAGYSGIGKSTL
    VOEIKPIITERGYFAGKFDQFOKNI PYSAVVAQOSLIRQLTENETQLQEMNKL
    LTVLGINAQVITIDVTPELILIGKQAVOELGPASQNRPNLVFNHPIHVPOKXHP
    AIFLDLQWADSTYTLKLDLIMRDKLGLFLIAGARDNENHNPITITIGLKSG
    LTIKELNLPMLNLTYNQUSISYISDFISAPPLAEVLSTKGSNPFVNOPLKTLTYQ
    EQLLI.FDFSSLWMHNI VOI BELGITDNVVDIMINKKL.PVNTQOILRLSACVNSF
    DLSTLAISEKSTFTSFDRDLPATQEGLLIPTSDLEITPADLVKSPILLINKFLDR
    VOQAAVALIDQKRNHILQIGOLLQONISETELNRIPENVDLNISWELLINIDOR
    VELSOELNAAKKAKEANAYSAALQYLISGINIPQDICEHYNALINAMHRESRPE
    ILGNPQOSLISILIEQAKTNLERKATINLLIVOTTLGNINQAIASRKRSLALGI
    YIPESDPEANLEVLINQANQLIANQDIAPINKPEVNIPEKIVAHHLNINIDPASTLA
    DOKLYKTI VLESNTLSLKYNGLGESYKGYTNQMLGALWGYQTYGEGYKALRKE
    KENNLAKQCKACFVLNQPOTNWKHKI KFRAINOEYHAGLESGLQFAGYILAEQLF
    NTL.FEGVLEDEVLAEDNGLEFVQRNOMQADMTLAARLGFNITSRDXDLGRIS
    EISEDEYLTGCEHRSEFAPLFCYLTFKSOILVLYDEDLAALQSLAEKLYLVYLGIL
    PVADONFVOSLILASTYLVKTPBEQAPFQKILNNQKMLKMANSPENPLIKITLVE
    AIAKISQYLEMLNLYDOAKISAQETFTONVALANELAKFWUSQGENIAKTYMI
    DAHYSYQLMGARKELDILAKY PQLGKTPWAGVNSLSTNHVDTTNRGSEFLDI
    ATVVKASQAIISGEIVLSELKIKLMSITINAGASNGCLLNREOLKIEVSINIEGSV
    CHSSTPSEDSPLSVINVARTEVDVLSAIDEPNFGNDLYIOSHSPKVLATVAT
    INOGKILGILYENNTIDPFTSDRELIIQILCSOASISLENALLIEOLERYSHNET
    KYBERTSQKAONKTIINQEKLASLGLIAGVAHERNPLNPNNAISAEKSEIL
    AIELNOSBTLNLSLEYIKGTTIDIVDNSVEIEQGLRADNITIQNNLLIARSDNNQ
    QFTDINSLETAVQALVHSLJNKLNGFNITIDYDQSLPEVAII.FODISRALINVD
    NACYAAYKKKQEVGDKFIPKLFVKTQNGESLIIIRDNGOGIAPEVIDIKINPEFTT
    KPPGEGTGLSLIHDIVVGOQGDIKINTGEVYEFVLTPIPNLSPEYQCKYD"
    complement(6345..6674)
    /gene="a112283"
    /complement(6345..6674)
    /gene="a112283"
    /note="ORF ID:a112283"
    /codon_start=1
    /transl_table=11
    /product="anti-sigma factor antagonist"
    /protein_id="BAB73982.1"
    /db_xref="GI:17131375"
    /translation="MYSNNTKIIOPSFGFDKGRQIHEESKITTSQVTTLLIDFO
    FTFMDSGSGFTLLTLKTVRQKKARLALCSINEQIKMLLELSDISKVEYVPPQAAF
    MGTVQY"
    complement(6809..7306)
    /gene="rbw"
    /complement(6809..7306)
    /gene="rbw"
    /note="ORF ID:a112284"
    /rbw gene product"
    /codon_start=1
    /transl_table=11
    /product="sigma-B activity negative regulator"
    /protein_id="BAB73983.1"
    /db_xref="GI:17131376"
    /translation="MSPHTDSCTHSHSTLOVASMDVAIVVMEFEKFNCKOLPYOIT
    EGOTLALIBGTNAVNRHAKHLNPTQVBLAVLIDPQYFOICIMDQSDINFRALAEKL
    SEKTDQAFNPLDHEAHWCIFPLKLKRDYAVMTVSYNRLAGANCLFLKKNLVPKVEL
    PNARD"

```

```

gene
CDS
    complement(7307..8308)
    /gene="a112285"
    /complement(7307..8308)
    /gene="a112285"
    /note="ORF ID:a112285"
    /codon_start=1
    /transl_table=11
    /product="glucosyltransferase"
    /protein_id="BAB73984.1"
    /db_xref="GI:17131377"
    /translation="WISPHTKPISLVYTRIHGTLVQLPPRLVYIEAISPREYQSW
    IEDSLTKIILDFQGTITIDSSGGSLSNKSQAKKIBELFNSIHPOVKLASLVG
    LEQILNIEANTDAIVPATRKIEQRPVTHPSVS PMKRAIDIVGAVIGLSVTANLFP
    PIAIAIKDSSPGVPLFSQTRCGMGRPRIRKFSVMYNAEADLQTTIPNOASGAFFKN
    DKDPRTIVGFEFLRTSLDEBPQPMNLTGMSLVGTRPTPNELEHYEVANMGRIDV
    KRGSGEYQVNGVRSKIRNREVDIQLDRLYQNMGLVYDLKILMTIIVFKKDSAV"
    complement(8305..8814)
    /gene="a112286"
    /complement(8305..8814)
    /gene="a112286"
    /note="ORF ID:a112286"
    /codon_start=1
    /transl_table=11
    /product="glucosyltransferase"
    /protein_id="BAB73985.1"
    /db_xref="GI:17131378"
    /translation="MILVTGTEQYPPNRLMSWIEVLQTELOBEIIVQYGNSTILP
    AGARYRPRKEKKEKQELHQAIRIVAHGEGTLLIDSLDPRYLVSQOQOEYVD
    HOVELGLASLSQINPVMSPGDLVRFIOAPROVAVDVSHSAIALCGLRRFRFQPL
    KELSGISK"
    complement(8811..9266)
    /gene="a112287"
    /complement(8811..9266)
    /gene="a112287"
    /note="ORF ID:a112287"
    /codon_start=1
    /transl_table=11
    /product="glucosyltransferase"
    /protein_id="BAB73986.1"
    /db_xref="GI:17131379"
    /translation="WKLVLICSGGHFGLQDLPFNMENHERVWTFRTATQAAITE
    EKVYWAYSPTRNRLPNLFRNLILAFQVLSKRPDLILSTGAVAAPFLILKLPFSQ
    TVFESVNTYITLSLSARVLPLSLVYQWPOQLTYQASLIVQESL"
    complement(9340..10371)
    /gene="a112288"
    /complement(9340..10371)
    /gene="a112288"
    /note="ORF ID:a112288"
    /codon_start=1
    /transl_table=11
    /product="glucosyltransferase"
    /protein_id="BAB73987.1"
    /db_xref="GI:17131380"
    /translation="WNLSEVILPVNSESSVAETLSVLAQTYRDLIIIVDGSYDKS
    IDICQFQDERIRIITHQONRGLAGARNIGIQASGEYLA FVSDULMPEKLAKLHE
    FERSPEVVSFRSRLIDDOCKPGIYOMPPLDITTPRYLCRANISNGSLIRRAV
    LDTIKFOENIAGEVDEPYRDSFQSGEIEICMLRIALQTTIKISIPALTLVYRNG
    GLSANVLYQYSWERILVKTQYINDEPIERNGNRKAYOLYILARRATRGSPAMAIV
    LHQAVKLHWRLILIEPHRTFISFAAYLLMLLPQFAYOPLERLVMITTSQEKRIQ
    KQILKPSI"
    complement(10457..11446)
    /gene="a112289"
    /complement(10457..11446)
    /gene="a112289"
    /note="ORF ID:a112289"
    /codon_start=1
    /transl_table=11
    /product="glucosyltransferase"
    /protein_id="BAB73988.1"
    /db_xref="GI:17131381"
    /translation="WTHASLISVITIPYNAKPLPAIASVEKQFTDWEILLINDG
    STDDTIEVVRHOKTSCDIYLINQNGVSAARQGVASHSGOIIAFPLADADQWL.PDK
    LSOHWQHQSRLGVSFAQVEIILQAGEPTGVSTSRINDLKEEYFLSENPPTTTSN

```



gene	2934.. 3953
/gene="cyt gene"	2934.. 3953
CDS	
/gene="cyt gene"	
/codon_start=1	
/transl_table=11	
/product="O-acetylserine sulphydrylase B"	
/protein_id="CA66130.1"	
/db_xref="GI:1296433"	
/db_xref="SWISS-PROT:Q44004"	
/translation="MVCNSIIIESIGNTPLVHRRLFGRADEAFKELINPAGSVKDRPAPVIVYKGLTEGALGHSHVIESSSGNTALALAMICGLGLRFTAVVPKISPTMLKILACGGEIVQVADKDSQSGYLETRIERLECOMREGDAMVINOVANLWESHYGGKAAIIAHVQAPADVIVLGVTTSGTVGIALRLKAMPMSVGVDAVPPCCANRRRPASTRALAPRAGAGIAVPERDRSPDPTRIHRHRGACMKRRRLRGSSGCAVVAITGVLHATVSHRPLRLTLFPDRGERIYLDIVYDAMLARIARTAVPNLAALPTVASLIRAGVAA"	
BASE COUNT	819 a 1496 c 1284 g 691 t
ORIGIN	
Alignment Scores:	
Prod. No.:	1.07e-18
Score:	381.50
Percent Similarity:	51.69%
Best Local Similarity:	32.21%
Query Match:	18.64%
DB:	1
US-09-889-746-2 (1-400) x AEMLEBQYM (1-4290)	
QY 27 ProaNGJuserLeu-----ProthrvAlGlnLeuHuproVAlValIle 41	
DB 717 CCATCCGATACGCTCCACGCCACGGCCACGCCACCACTGCTGGCCCCGGGTGAAGTC 776	
QY 42 ThrIleAspLySerGlywMetAlaLeuAlaAaArgIleThrGlnMetProHlsthThr 61	
DB 777 ACCGGGTACGCAACATCGAGCATCTGCTGCCCGCCGACCGGCA-----TCG 824	
QY 62 LysValIleTyrGlnGlnGlnIleGlnGlnGlnAlaIleThrGlySerArgIleuAlaAsp 81	
DB 825 AC-GTGTGGTGAAGCACCAGCACTCGAGAACTGAAACCGGGCTCGGACAGCTGGCCAG 883	
QY 82 ValMetAlaGlnLeuIleProSerLeuGlyValSerSerGlyThrThrSerAsnProGly 101	
DB 884 GTGCTGTCCAAAGTATCCCGGCAATGCGGCACTCCAGCACACGGTGCACAGACTTCGGC 943	
QY 102 GlnTrMetHlEgIYArgGlnValGlnPheLeuLeuAsnGlyValProLeuThrGlySer 121	
DB 944 CAGACACTCGCGGGGCGCGGCATGCTGTCGCTGCTGCAAGGCATTCCTGTAACACCAAT 1003	
QY 122 ArgAspIleSerArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGluVal 141	
DB 1004 CCAAGACTCGCGCGCAATCTGGCCAAATTCGATCCGGCACTGGTGAAGCAAGTGAAGTG 1063	
QY 142 LeuSerGlyValIleThrSerIleTyrGlySerGlyAlaIleThrGlyGlyLeuIleAsnIleVal 161	
DB 1064 CTGCGAGGACAGTACCCGATCTACCGCGCGCGGCCACCAAGCGGCGGCATCTGTCACATCACCC 1123	
QY 162 ThrLysSerAspLeuGlnGlnGlnPheGlnIleThrArgIleGlyValAlaHlsgly----- 179	
DB 1124 ACCCGGCGACCGCGCGGCGCAAGCCCAAGCCGATACCAAGTACGATGCCACCGCGCTC 1183	
QY 180 SerLysLeu-SerSerGlnGlyIleGlyTyrGlnValGlyGlnSerValAlaGlyValSer 199	
DB 1184 TCCCGCTCGCGCGGAGCGGCTTGAAGCGCAAGTGCAGACGACACTTTTCCGAGCAGAG 1243	
QY 199 rGluGlnGlyAsnValAlaLeuAlaArgLeuAspValaAspTyrArgThrThrGlyGlyAlaPh 219	
DB 1244 C-----GGAGTGTCTGACTATGCTTTCACCTTACAGCGCGCGCCATATCGGCAA-TCGTA 1296	
QY 219 eAspAlaAsnGlyLysArgIleAlaProGluProAlaGlnIleThrAspLysGlnAspSerLys 239	
DB 1297 TGAATGCAAGGGCGCA-CGCATTGCGCGGAGGCCAAGCAGGGCGACCTGTTGCATCTGAA 1355	

Oy 239 sSerIuSerValenThrAenValaSpTyrGlnLeuAbaSpAspYsgInaenIleAenIle 259  
 ::::: ||||| :::::  
 Db 1356 CAACTACAACTCGCGCGCAAGCTCGCGCTCGCTTCGCGCCGACCGACCGCATCCAGCTAG 1415  
 Oy 259 uAlaLeuThrIstYrAenAbaSpYsgInaSpThrAbaPyrAlaProAbaPyrGlyAbaNa 279  
 ::::: ||||| :::::  
 Db 1416 CCAAGTTCAGCATTTTCAGCGCCAGCAGAGACACGCGCTCGCGCTCGCAT-----CCGAG 1469  
 Oy 279 gLeuAlaValLeu---PheGlyGlyLysPProSerIeuaSnalIleLysGlyLeuSerIe 298  
 ::::: ||||| :::::  
 Db 1470 CGTGCCAAACTGCGCGCGGACGTCGCCGCG-CCGCGCTTCGACGGCTGCAACT 1527  
 Oy 298 uSerGlnGlnProLysThrThrLysSerThrPheasnIleAenTYrHisAspAspLe 318  
 ::::: ||||| :::::  
 Db 1528 TGCCGACCAAGAACCAAGATCCGCAACAGCTCGTCAATTCGAATTCGAACACCGGACCT 1587  
 Oy 318 uTPGlyAbaThrIleAenThrAbaSnalAbaTYrTYrArgArgGlyLysGlyAbaPheTYrPr 338  
 ::::: ||||| :::::  
 Db 1588 GCTCGCGACCGCATGTGTGCGCGCATTTCTACTACCGGAACACTACACCGCTTCACCGC 1647  
 Oy 338 cPheValAlaProPheSerIleAlaValAlaLeuProIleLeuGlnSerMetAbaLeuPr 358  
 ||||| ::::: ||||| :::::  
 Db 1648 GTTC-----GACGCGCGCGCGCTGCGCTG----- 1672  
 Oy 358 cSerAlaThrLeuAbaPyrAlaTYrThrLysAlaProGlnAlaArgAlaTYrGlyValLeuG 378  
 1673 -----CGCGCGCGCAACGTGACCA 1692  
 Oy 378 nSerGlnSerLysAlaGlnValLeuGly-----ArgValProAbaLeuAs 393  
 ::::: ||||| :::::  
 Db 1693 GCGCATGACGAGATTCGAGGCTTCGCGACGCGCTCGACCTTCAGACGCGG---CTCGG 1749  
 Oy 393 nLysProLysArg 397  
 ::::: ||||| 1762  
 Db 1750 CCGGACAAAGCGC 1762

RESULT 25  
 AP003590/c 333500 bp DNA linear BCT 28-NOV-2001  
 LOCUS  
 DEFINITION Nostoc sp. PCC 7120 DNA, complete genome, section 10/19.  
 ACCESSION AP003590 BA000019  
 VERSION AP003590.1 GI:17131676  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Nostoc sp. PCC 7120 DNA.  
 Nostoc sp. PCC 7120  
 Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 1  
 Kaneko, T., Nakamura, Y., Wolk, C.P., Kuritz, T., Sasamoto, S.,  
 Watanabe, A., Iriguchi, M., Ishikawa, A., Kawashima, K., Kimura, T.,  
 Kishida, Y., Kohara, M., Matsumoto, M., Matsuno, A., Muraki, A.,  
 Nakazaki, N., Shimpo, S., Sugimoto, M., Takazawa, M., Yamada, M.,  
 Yaezaki, M. and Tabata, S.  
 Complete genomic sequence of the filamentous nitrogen-fixing  
 cyanobacterium Anabaena sp. strain PCC 7120  
 DNA Res. 8 (5), 205-213 (2001)  
 2 (bases 1 to 333500)  
 Kaneko, T.  
 Direct Submission  
 Submitted (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research  
 Institute, The First Laboratory for Plant Gene Research, Yana  
 1532-3, Kisarazu, Chiba 292-0812, Japan  
 (E-mail:kaneko@kazusa.or.jp,  
 URL:http://www.kazusa.or.jp/cyanobase/  
 Tel:81-438-52-3935(ex.2338), Fax:81-438-52-3934)

FEATURES  
 source  
 1..333500  
 /organism="Nostoc sp. PCC 7120"  
 /db\_xref="taxon:103690"  
 /note="synonym:Anabaena sp. PCC7120"  
 complement(57..926)  
 /gene="a112584"

gene





```

/gene="all2020"
/complement (761. .3619)
/gene="all2020"
/notes="ORF ID:all2020"
/codon_start=1
/transl_table=1
/product="ABC transporter ATP-binding protein"
/protein_id="BAB73721.1"
/db_xref="GI:1713111"
/translation="MKLQGTYYINGNEPIILLNDFSRVYVVGSGSMALFAVTVKDVGI
DGRYLCISOGALFTVQOQIILAVPIGEVELLLHPECFRGLCAKANAVIENW
LMEISVLCOIPTKI OVKEGQERFSLNNGQNPQPAAGLTCTWQLQEGNCFIAPGER
LILATVAFPNQDKWMLTPVCGVQLTHTTSEYENSIDLLAGLSQTLTQILSCVTLID
EQEAQAEILQDERLNRRTVTAELGSLNSQDQPFLEGTPLPLVAAGVGRAM
GIKISPPRSNLQVKEPELEAVRASIRMRVLLRDNWMLDAGALVATYKSNQPV
ALLPAGNRYEYFDPNYRVVNVGCVATLAPVAYMFYSPDRAALRAIDIVKFTLRG
HMDLLOLFTGTFITTLGMITPVATSIIDNAPDSRSLLOIGLGLVVAALGAGL
PQAGGALLAMETSGDASTOAGWDRLLNLPVFFREYTTGDLSSISSVAIRROL
SGRLINPITSIPALFYQLFYNNYKALIAVASAVVAIAITISIGILLSSKVRPLL
EVQGNIFQTVQLINGISKHIAGAERFAAWSKNYTRQIKJELSTOLVEDFVIFN
TIIPITSGVLFWFTKLLLEBAQTSIAFSISLGTFLAFNATFSNFTGTNLNSTVQ
ILOQLPQWRQPIITPIPEVNLKSQDPGLAGKITVDNVTFCYRRDGNLTLDNVSIS
AEPGFALVSGSGKSTLLRLLLGPETQAGSIHYDQDLSGLDVAVRQGLVGL
ONGQNSAIFENIAGDAQITLDREAWAARMGFAEDVTAMPQMOMHTVVSQGGNLSG
CQRORLLIARALAKPRILLFDEATSALDNRTQAI VSES LDKLVTRIAIAHRLSTIR
NAHRIYVQAGRVVQGGTFQELANTEGIFAQIMARQMA"
/complement (3819. .5957)
/gene="all2021"
/complement (3819. .5957)
/gene="all2021"
/notes="ORF ID:all2021"
/codon_start=1
/transl_table=1
/product="ABC transporter ATP-binding protein"
/protein_id="BAB73720.1"
/db_xref="GI:1713112"
/translation="MKPTLLQMEAVECGAAALGIILGYGRVPLPOLRRDCGVS RD
GSKASNMVKAARTYGLEAKGPKKELKQVQQLRPPIYVPMNPHFVVEGFGKHQVYIN
DPATGPRHVSQDEFDEGTYGIVLVMEPGAEPKQGRKPSMTIASLMSLQGAESALVYC
IVAGGFTIIGLVIPFVSQI FVDVILVQGRMLWHLPLALLAMIAI AILQSGISTLLRLRY
LRRLKIRLAVQMSRFLWHILRLPVSPAQRVAFAGEISNRTLNIDANQRLIOEYKKAAG
IDTMLVPLVAMQYDMLVRLTMVVSFAA NVLTLQMSRORIDANQRLIOEYKKAAG
ISIALQSIETLAKAGLESDFSRWSGYTHALNSQBELGINTQTSFVLPILLALSAL
MLLIIVGGRVMDGHSILGMLIAFOGLMSGFQQPVNMLVNFGLTLEGLNLRDDV
LDNPFVEGRTSSSSQALPKQLOQYVELKNITFGYSRLEAPLIENFNLSIKPGORVA
LVGSGSGKSTIAKLGLYQVWTGCEIYFDSPREEISLQLLTNSVAMVEQEIILLFGG
TVRNLTWDTTVPDKNLMRACQDAAIADVIFSMSSGGYDAELIEGAANLSGGQQRLE
IARALVNPSILVWDEATSALDAETEKIDQNLRRRGCTCIIVAHRLSTIRDCDQIIV
LERKVVOQGTQHWELQVEGAYSRLIRTEG"
/complement (6028. .7395)
/gene="all2022"
/complement (6028. .7395)
/gene="all2022"
/notes="ORF ID:all2022"
/unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAB73721.1"
/db_xref="GI:1713113"
/translation="WQVSPKSMPLVALGSI VGVAFIWSIYGSIPITVEGRGVLIYP
SNVPLQSKSAGQIMALNVKGVDIKKQVLAITDQAEKRLQLOQOGRKLTQLSQQX
AVGSLQRRLEOBERSOQOQRYURKILELEAITDPLTKNTGNSVDQOQOGLQRLQ
QAQALTAVALRMLKIRBELFQKGAISGDEAKABOEYLNKLEIKIAQAKELDVK
ETQOQKRENLTLVADLQSLQKLDQSOASVADQENLVARNKEIQEVKREIGSLE
LQLSNSQIISQHSRIILEITLTPGVVNACTRLATIEAENPSQKLGVGVTFPVAEGK
KIQPMNSIOTPTQVKEBFQGI VGNVTSISRFITKEAANAENGVNSVLEGLVSQQQ
GLIQVFSHLELDANTPQSKSSSTGPHLNISSTGTTTVRVKVERAPITTVLFI LRS
ASGIY"
/complement (7472. .9502)
/gene="all2023"
/complement (7472. .9502)
/gene="all2023"
/notes="ORF ID:all2023"

hypothetical protein"
/codon_start=1
/transl_table=1
/protein_id="BAB73722.1"
/db_xref="GI:1713114"
/translation="MSDCLVLA MPYAAVERPSTALGLLKARLOQAKLEPTILYENIOP
AESIGLDVYAMISENFIETELGKTSVGLFPFEPNHSSEYFDLISFCFKSYENLYPD
RNLEKPLAVREGATAPIESVAQSVANQRIIGCSSTFOOHCASLALKKRVRELAPE
IITFMGCAGCEGATLTHFSWDFIISGEGDILNFECYQLLDGGRVDPISILPY
GVSPSHRTSKISIAAPRASVQNLQDQPIPDYDDVFQTLHASTLSYIOPGLLIETS
RGWQGHKHTFCGLNGSINYSKSPDLVKFTQLCERYGLRKFQVVDNILDMMNH
MNTVIVPVLAGEAYNI FVETKANLWEHLQOLAQAGVRCIOPGIESMHSILKMNK
CNTAMNVQLLKWARLGIIVFNILADPFGESEDEWYEMMELPLAHLOPPSGASE
IRDRFSPYTERPSDMLNSPLRPYKIVVYLAPEILSELAYFEDQNNLPVRYTAE
NEGTSEKTERPGLLENKLVNQNWLQSYLYKEHKPEYLPILSMNDSGLSITTYDSR
PCATSEILFLEGLAYVYVYCDRALTYREVVSINKKLDVPVSWDEIFPTVSELREKR
ILLIDORLLSALAKGSCSPYLSVSELPAGYIDLHBEPRRHNSKSEKVLPLPN"
/complement (9616. .9909)
/gene="all2024"
/complement (9616. .9909)
/gene="all2024"
/notes="ORF ID:all2024"
/unknown protein"
/codon_start=1
/transl_table=1
/protein_id="BAB73723.1"
/db_xref="GI:1713115"
/translation="MSLVMTWKELEKIIIVKAGODESKOALVSNPRSTLEKEGINLP
SSIDVNLVETTPGNLSLQALPNSEODT LSEAELESVAGGISGEVSVTVVKVF"
/complement (10195. .13482)
/gene="all2025"
/complement (10195. .13482)
/gene="all2025"
/notes="ORF ID:all2025"
/similar to MreD protein"
/codon_start=1
/transl_table=1
/protein_id="BAB73724.1"
/db_xref="GI:1713116"
/translation="MNVSKQLIKI ITERSNI SERLSSESIFNDRIQINANNKALEK
WYOFIAGKGNHNFEXKRLLDGDLTNTISCVLGNVNLADKELVPNWETLAKGQVATP
VPIENSNKLOQHRVLDPRPFPFEELVPVDVAKKLIKESAGSSYLLSQMVAHIQ
MERSLLEHLGIGICISPLESEFSIERSLSKSELVRFINQLOSNCSKQYTFVKNLLKD
GLASFPOEYSVLARLATSVDVFWATGEFSLASDSKIQLTFFQEKLGQVSVQ
LANSDFNRRMGLOQRAVITVPASGPKLVYKPKSLGLEKAYFDFLDWINQKINLPLKLLKI
INCSNYGMEFAFALPCQODADAVHYVQRAGMVLICVYLLKGNDCHECENIACGDQV
LVDELTLHRTWLSKDDAKSITANECLQDSVVGTGFLPWQIILPYEOTEILLKDFS
GLGFGCEQMPYRMAWKHINTDSMVI VREYTKLLPKNRPGENIDTSLNHSKELL
DGRQMTQFLOQRKEELLASDSPITAFSNKRVLRVINTTVIASILONSNNHKCLRTG
VHESIQDLSRAPLSESDKHLPLWLLAAEKQALEQDIPYFTAYSQNSNVDISSEOT
INKFLSSSYDDVIAHLRQMDADLAKQISLIRGSYSLCTNEQLNSLPENPNLYMD
IVAPIERRGMLOQAIAIAQOLEQOAIRKDNSVTWIGMVLFEARFQLOPVGNSLYD
GCCISLPLAALASVTNRAEFGELSLASLQRLTQDPI DLQOKLIGOMGIGATG
LASIVVALVPISEFLGEVELI QDAQOLASLVINEATVNOAPGVTGAGTILSLLAL
KYATKEPATLOQAISWGNLYLDRVTVDEVIQAWTNAEGVLLGFGSHGAGIATLLO
LYAATDEPIFASAREAITLEQRLFSLSNCCQDLDLVSDNLSLSTTESCSGLIGIS
VARLGTNLIDTDEIRQELINLHTSHSSLOVDNITCCGNFSEHMLLVASKQLSHS
ELLETTANNRATQVLSRANQIGYFOLLPHLPADVVSFGFCQGIAGIGYELLRLSPDLL
PSILLWQ"
/complement (13877. .14227)
/gene="all2026"
/complement (13877. .14227)
/gene="all2026"
/notes="ORF ID:all2026"
/similar to nitrate hydratase"
/codon_start=1
/transl_table=1
/protein_id="BAB73725.1"
/db_xref="GI:1713117"
/translation="MSBOTKTRKEIEAQIIVQAMKDETYRQELNNKSNKAVIERFAIQ
LPBEINVVHVVEENDSNFYFIPARNPLENDELSEQLEAVAGSGLDIFTFVETKYDR
LKQFGKEIYKAIYD"
/complement (14385. .14720)

```

CDS	/gene="a112027"
	/complement(14385..14720)
	/gene="a112027"
	/note="ORF_ID:a112027
	similar to nitrite hydratase"
	/codon_start=1
	/transl_table=1
	/protein_id="BAB73726.1"
	/db_xref="GI:17131118"
	/translation="MSGSQTRKDVKEAQIVQANKDAEYRQELLNPKKIIVEDFGVQV
	LPEENITVAHMEENSNTLFFVIIPARNLEDELSDOLEAVAGSLMTLLIPIAIG
	AIEHNNSRK"
gene	/complement(14854..15042)
	/gene="a812028"
	/complement(14854..15042)
	/gene="a812028"
	/note="ORF_ID:a812028
	unknown protein"
	/codon_start=1
	/transl_table=1
	/protein_id="BAB73727.1"
CDS	
Alignment Scores:	
Pred. No.:	2,93e-12
Score:	326.50
Percent Similarity:	46.18%
Best Local Similarity:	27.83%
Query Match:	15.95%
DB:	Gaps: 11
US-09-889-746-2 (1-400) x AP003588 (1-334520)	
Oy	27 ProbenGluerleupProthrValGluLeugInuProvalValIlleThrIleaPlySer 46     ::: Db 235532 CCTACTCCACACCACCGCAAGGTAAACCCCATCGGAACAGCACAAGCATGTGCAT 235591     ::: Oy 47 Gly-----MetAlaLeuaIlaenArg-----IlleThrcInmetPro 58       ::: Db 235552 GACCCGATCGAGTGGTGCTGTCAGCGCACMAAGGGAGAGCATATVCAAATGTACCG 235651     ::: Oy 59 HisThrThrLyValIlleTyrgLuJugInillegIngluAlaThrGlyserArgIn 78     ::: Db 235652 CGATCGGTGACGGTGAATTACCCGTGACCACTGCCAACGAACTACTTTATCGTAT 235711     ::: Oy 79 LeuAlaAspValMeValagInleuleProSerLeuglyValSerSerglyThrThrSer 98     ::::: Db 235712 TTAACCTCGATCTGTTGATAACAGTCTCTGGTTGGGTGCTAGTGGCAATCACAAACA 235771     ::::: Oy 99 AsnPhelgylInthechi seGlyatrglnValGlnPheluLeuabngilyalProleu 118     ::::: Db 235772 AGTTTGTCCACACTTGGCGGGAGAACCACCTTGATCTTGATGGATGGATACCCAAT 235831     ::::: Oy 119 ThrGlySerArgAspIleSer-----AcgGlnLeuasnSerIleasnProasngIn 135     ::::: Db 235832 AGTTCARATATGATATATATATATATCARGCGTTCCTAACTTAGCGGATTCATGTGGGCG 235891     ::: Oy 136 ValAlaArgileGluValLeuSerGlyalaThrSerIleTyrglySerglyalaThrGly 155     ::: Db 235892 ATTTGAACGATCGAAGTGTGTAGAGGGCCGAGTGGCTTTATGGTGATGTCTCGCGGT 235951     ::: Oy 156 GlyLeuIleAsnilleValThr--LyseSerApbleuGluGluGlnPhelguThrarg 174     ::: Db 235952 GGGGGGATTAATAATATATATCCCGAAGACCTGACCAAGAGAGTGTTCACACGAGNG 236011     ::: Oy 175 IleGlyValHisGly---SerLySeuSerSergluGlyIleGlyTyrcInValGlyIn 193         Db 236012 ATTGGATATCGTTCGTTGGTGAATCTCAATCTTGGAATTTTTGGTNACCTTTGTTAATTAC 236071         Oy 194 SerValAlagiValaserGluabnglyaenValLeuAlaArgLeuaspValaPyrarg 213     ::: Db 236072 GGATATTTCTGTCAGCAAGAGGGGGAGATTTTATGCGATCTCTTAACCGCGAT----- 236125         Oy 214 ThrThrGlyGlyAlaIleAspAlaengilylsrGllleAlaIpriGluProAlaGlnThr 233

Db 236126 AGCTTTGAAACCCCTTGATGCGAAGGCATGCAATT-----CCATTATTGGT 236176  
 Oy 234 Asplysgln--AppSerLysSerLeuSerValAsnThrAsnValAspTyrGlnLeuAsp 252  
 Db 236177 GACGACGACGAAATATGAGTCTTCAATTACGTTTAGGAAATATGGGTTTCAATTGGGT 236236  
 Oy 253 AsplysglnAsnIleAsnLeuAlaLeuThrHisTyrAsnAsplysglnAspThrAspTyr 272  
 Db 236237 TCGGAAACGAGATTGCAAATTATACAGCCAAATTATATGATGATGATCAAGTAATGATGTA 236296  
 Oy 273 AlaProAspTyrGlyAsnArgLeuAlaValLeuPheGlyGlyLysProSerLeuAsnAla 292  
 Db 236297 -----GATTATGAT-----TTACCTAGACAAATACAGCAATCAAAAA 236338  
 Oy 293 IleIysGlyLeuSerLeu-----SerGlnGlnProLysThrThrLys 306  
 Db 236339 GCACGAGCTTTAGATTCACACGATGGAATTATCAATCTTACAAATCCTTTAACCGAGA 236398  
 Oy 307 SerThrPheAsnIleAsnTyrHisIleAspAspLeuTyrGlyAsnThrIleAsnThrAsn 326  
 Db 236399 ACAGTAATTCAACTGATTAACCCACATATATCTGATAGTCAACTACAAAGCCGAG 236458  
 Oy 327 AlaTyrTyrArgArgGlyLys 333  
 Db 236459 GCTTATATGCTCAACTATAA 236479

RESULT 27  
 LMFLCHR36\_27/c  
 MPOCOMMENT

Sequence split into 36 fragments LOCUS LMFLCHR36 Accession AL499624

Fragment Name	Begin	End
LMFLCHR36_00	1	110000
LMFLCHR36_01	100001	210000
LMFLCHR36_02	200001	310000
LMFLCHR36_03	300001	410000
LMFLCHR36_04	400001	510000
LMFLCHR36_05	500001	610000
LMFLCHR36_06	600001	710000
LMFLCHR36_07	700001	810000
LMFLCHR36_08	800001	910000
LMFLCHR36_09	900001	1010000
LMFLCHR36_10	1000001	1110000
LMFLCHR36_11	1100001	1210000
LMFLCHR36_12	1200001	1310000
LMFLCHR36_13	1300001	1410000
LMFLCHR36_14	1400001	1510000
LMFLCHR36_15	1500001	1610000
LMFLCHR36_16	1600001	1710000
LMFLCHR36_17	1700001	1810000
LMFLCHR36_18	1800001	1910000
LMFLCHR36_19	1900001	2010000
LMFLCHR36_20	2000001	2110000
LMFLCHR36_21	2100001	2210000
LMFLCHR36_22	2200001	2310000
LMFLCHR36_23	2300001	2410000
LMFLCHR36_24	2400001	2510000
LMFLCHR36_25	2500001	2610000
LMFLCHR36_26	2600001	2710000
LMFLCHR36_27	2700001	2810000
LMFLCHR36_28	2800001	2910000
LMFLCHR36_29	2900001	3010000
LMFLCHR36_30	3000001	3110000
LMFLCHR36_31	3100001	3210000
LMFLCHR36_32	3200001	3310000
LMFLCHR36_33	3300001	3410000
LMFLCHR36_34	3400001	3510000
LMFLCHR36_35	3500001	3529852

Continuation (28 of 36) of LMFLCHR36 from base 2700001 (AL499624 Leishmania major chromosome)

Alignment Scores:  
 Pred. No.: 4.63e-11 Length: 110000  
 Score: 302.00 Matches: 70

Percent Similarity:	60.99%	Conservative:	41
Best Local Similarity:	38.46%	Mismatches:	63
Query Match:	14.75%	Indels:	8
DB:	2	Gaps:	3

US-09-889-746-2 (1-400) X LMFCHR36 27 (1-110000)

Qy	12	LeuProLeuLeuSerValAlaValThrGlnGlnLeuTyrAlaGlnProAsnGlnSerLeu	31
Db	96012	GTCCCGCTTTCCACGCTGGCG-----CAGCAGGCGCGCTCGCGCCCCCGGACGCGCG	95959
Qy	32	ProThrValGluLeuGluProValValIleThrIleAspLysSerGlyMetAlaLeuAla	51
Db	95958	CCCGCGCATGAGCTGAGCAGCGTCCAGGTGGGTGCGCCGGATCAGAGCGATCTGGTC	95899
Qy	52	AsnArgIleThrGlnMetProHisThrLysValIleTyrGluGluGlnIleGlnGlu	71
Db	95898	AATCCGACGCGCAGGTC-----ACCGTGATCGAGCGCGAGCAGCATCGAGGAG	95851
Qy	72	GlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGly	91
Db	95850	CTGAAGACCGGCTCGACCAATCTGGCAGCATGTGTAGCAAGCTGATTCGGCGCATGCC	95791
Qy	92	ValSerSerGlyThrThrSerAsnPheGlyGlnThrMetHisGlyArgGlnValGlnPhe	111
Db	95790	GATTTCGAGCCGCAACCGTGACCGATTCGGCCAGACGCTCGCGGACGCGCGCTGGTG	95731
Qy	112	LeuLeuAsnGlyValProLeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIle	131
Db	95730	CTGGTCGACGCGCATTCGCTCAATACCAACCGCGATTCAGTCTCGCAATCTGGCAACGATC	95671
Qy	132	AsnProAsnGlnValAlaArgIleGluValLeuSerGlyAlaThrSerIleTyrGlySer	151
Db	95670	GATCCGACCAACATCTGAGCGGATCGAGTGTGCGCGGACAGCGCGCTCTACGGCGCC	95611
Qy	152	GlyAlaThrGlyGlyLeuIleAsnIleValThrLysSerAspLeuGluGluGlnPhe	171
Db	95610	GGCCGCGAGCGCGCATCATCTCGATCAGCACCCGCGACGCGGCGCGGCCACCGCC	95551
Qy	172	GluThrArgIle-----GlyValHisGlySerLysLeuSerSerGluGlyIleGlyTyr	189
Db	95550	GACACCAAGGTAAACGCTGGGCACCGCGTGTAGCGGGTCTACCAAGAGGGGCTCAGCGCC	95491
Qy	190	GlnVal	191
Db	95490	AACGTG	95485

RESULT 28  
LMFLCHR32\_10  
WPCOMMENT

Sequence split into 28 fragments LOCUS LMFLCHR32 Accession AL499622

Fragment Name	Begin	End
LMFLCHR32_00	1	110000
LMFLCHR32_01	100001	210000
LMFLCHR32_02	200001	310000
LMFLCHR32_03	300001	410000
LMFLCHR32_04	400001	510000
LMFLCHR32_05	500001	610000
LMFLCHR32_06	600001	710000
LMFLCHR32_07	700001	810000
LMFLCHR32_08	800001	910000
LMFLCHR32_09	900001	1010000
LMFLCHR32_10	1000001	1110000
LMFLCHR32_11	1100001	1210000
LMFLCHR32_12	1200001	1310000
LMFLCHR32_13	1300001	1410000
LMFLCHR32_14	1400001	1510000
LMFLCHR32_15	1500001	1610000
LMFLCHR32_16	1600001	1710000
LMFLCHR32_17	1700001	1810000
LMFLCHR32_18	1800001	1910000
LMFLCHR32_19	1900001	2010000
LMFLCHR32_20	2000001	2110000

LMFLCHR32_21	2100001	2210000
LMFLCHR32_22	2200001	2310000
LMFLCHR32_23	2300001	2410000
LMFLCHR32_24	2400001	2510000
LMFLCHR32_25	2500001	2610000
LMFLCHR32_26	2600001	2710000
LMFLCHR32_27	2700001	2727709

```

US-09-889-746-2 (1-400) x LMFLCHR32 10 (1-110000)
Alignment Scores:
Pred. No.: 1.06e-08 Length: 110000
Score: 270.00 Matches: 60
Percent Similarity: 54.39% Conservative: 33
Best Local Similarity: 35.09% Mismatches: 66
Query Match: 13.19% Indels: 12
DB: 2 Gaps: 3
Continuation (11 of 28) of LMFLCHR32 from base 1000001 (AL499622 Leishmania major chromosome 2700001 2727703)

```

QY	175	IleGlyValHisGlySerLysLeuSerSerGluGlyIleGlyTyrGlnValGlnSer	194
Db	82389	CTGGGACGCGCTGACGCGGCTCAGCAAGGAGGGGCTCAGCGCAACGTCGACGACGAC	82448
QY	195	ValAaglyValSerGluAanGlyAsnValLeuAalArgLeuAspValAspTyrArgThr	214
Db	82449	TTCTCCGGC-----AGCCAGGAGCCGATCGACTACGCTTCACCGTCGCGCGCAGCGC	82502
QY	215	ThrGlyGlyAlaPheAspAlaAsnGlyLyAArgIleAlaProGluPurAlaGlnThrAsp	234
Db	82503	ATCGGACGCCCTTACGACGCGCAGCGCCATCGGATCGCGCCGAGCCGACGACGGCGAT	82562
QY	235	LysGlnAspSerLysSerValAsnThrAsnValAspTyrGlnLeuAspAspLys	254
Db	82563	CTGTTCGATTCCAAGCTTACAACATCAGCGCAAGCTGGGCTGGCGCATCGACGGCCAG	82622
QY	255	GlnAsnIleAsnLeuAlaLeuThrHisTyrAsnAspLysGlnAspThrAspTyrAlaPro	274
Db	82623	CAGCGGCTCCAGCTGAGCGCTTATCAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGG	82682
QY	275	AspTyrGlyAsnArgLeuAlaValLeuPheGlyGluLysPro-----SerLeu	290
Db	82683	GACCCGTCG-----GTCCGAGCGCTGCGCGCGGCTCGGCGACGGCG	82724
QY	291	AsnAlaIleLysGlyLeuSerLeuSerGluGlnProLysThrLysSerThrPheAsn	310
Db	82725	CGCGGATCCGCGGCTGACGCTGGAACACCAACGACGCTGGCAACACGCGTCTGAC	82784
QY	311	IleAsnTyrHisAspAspLeuTyrGlyAsnThrIleAsnThrAsnAlaTyrTyrArg	330
Db	82785	CTGGATTACCAACGACGAGGACTTATACGCGCAGCAGGTATCGACGCGATTTCTACTATCG	82844
QY	331	ArgGluLysGlyArgPheTyrProPheValAla	341
Db	82845	GACTACTTCACGGGCTTCTCCGTTTCGACGCG	82877
RESULT	29		
LOCUS	AE012011/c		
DEFINITION	AE012011	12514 bp	DNA linear BCT 29-MAY-2002
ACCESSION	Xanthomonas axonopodis pv. citri str. 306		section 389 of 469 of
VERSION	AE012011		the complete genome.
KEYWORDS	AE012011.1	GI:21109982	
SOURCE	Xanthomonas axonopodis pv. citri str. 306.		
ORGANISM	Xanthomonas axonopodis pv. citri str. 306		
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;		
AUTHORS	Xanthomonas		
	1 (bases 1 to 12514)		
	da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R.,		
	Queiroz, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida		
	Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,		
	Camargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J.,		
	Chambergo, F., Cianella, L.P., Cicarelli, R.M.B., Coutinho, L.L.,		



TITLE	Comparison of the genomes of two <i>Xanthomonas</i> pathogens with differing host specificities
JOURNAL	Nature 417 (6887), 459-463 (2002)
MEDLINE	22022145
PUBMED	12024217
REFERENCE	2 (bases 1 to 12514)
AUTHORS	da Silva,A.C.R., Ferro,J.A., Retnach,F.C., Parah,C.S., Furlan,L.R., Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Camargo,L.R.A., Camarotte,G., Canavan,F., Cardoso,J., Chamego,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L., Cursino-Santos,J.R., El-Dorri,H., Faria,J.B., Ferreira,A.J.S., Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A., Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira,J.R., H.A., Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinoia,L.A.F., Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and Kitzime,J.P.
TITLE	Direct Submission
JOURNAL	Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de São Paulo, Av. Prof. Lineu Prestes 748, São Paulo, SP 05508-900, Brazil
FEATURES	Location/Qualifiers
source	1. 12514
gene	/organism="Xanthomonas axonopodis pv. citri str. 306"
CDS	/strain="306"
gene	/db_xref="taxon:190486"
CDS	/note="pathovar: citri"
gene	complement(150. .368)
CDS	/gene="XAC3604"
gene	complement(150. .368)
CDS	/gene="XAC3604"
gene	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
CDS	/codon_start=1
gene	/transl_table=11
CDS	/product="conserved hypothetical protein"
gene	/protein_id="AA038447.1"
CDS	/db_xref="GI:21109983"
gene	/translation="MFEQPTEDIALMKSDPQKLYORHLKNCMDAEGVLPIDVTLGOMKREKLAQKCLTRWYDGLTH"
CDS	complement(720. .1610)
gene	/gene="upb"
CDS	/note="XAC3605"
gene	complement(720. .1610)
CDS	/gene="upb"
gene	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
CDS	/codon_start=1
gene	/transl_table=11
CDS	/product="outer membrane protein"
gene	/protein_id="AA038448.1"
CDS	/db_xref="GI:21109984"
gene	/translation="MKLRPLLYGVSLVTLPLAAKMDLQADALNRRLAONDPQNDLAFERLQGSVALAKRRDRDELVPLARRVRIEITRTYALRELRLEGRNDLLEGRDARARQEARLRVQIOTBAQROAEQALADQAEALRTSVAGKQTAALAAQKAVQLAHEAEVLVAGALPAKPDSEGEVSLSGAFGGKAA LSGDAAGAKALAEYINIGKGRVSVIGFTDPAIAIKRAEALRDALVAGVSAHLQ
gene	VSQTKGASKARAAEVVAP"
CDS	complement(1607. .1990)
gene	/gene="upb"
CDS	/note="XAC3606"
gene	complement(1607. .1990)
CDS	/gene="upb"
gene	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
CDS	/codon_start=1
gene	/transl_table=11
CDS	/product="outer membran protein"
gene	/protein_id="AA038449.1"
CDS	/db_xref="GI:21109985"
gene	/translation="WTASPAVYLRSLYGIACFMVLSPPAAQVAPELTAACAVORATQADQVAPDLVNLARQELMQAQAQKQKQKQVQJALAAADADLAKARSEAAVTAQLQRKQVAYQLQNTINTGEGR"
CDS	2089. .3222
gene	/gene="upb"
CDS	/note="XAC3607"
gene	2089. .3222
CDS	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
CDS	/codon_start=1
gene	/transl_table=11
CDS	/product="type II secretion system protein-like protein"
gene	/protein_id="AA038450.1"
CDS	/db_xref="GI:21109986"
gene	/translation="MDIGYFLKMTKRNASDFLTGAPVYIKIEGKIYPLQNTGLPQVAKKIASLDEQGVQPEBELNMAIALPDACRFVNVFKQGEVGMVYIRAIRSRIPSIELMLPOLKXVIMTPRGLVYVASTGSKSTSLASMDHNSYTTGHILIEDPIEYHKRMSIVNOREVGLDPHAFQSAKXNMRAPVYILGETILDATMEALAPAETGHCIATLSNNAQDRIERLINPEPSAKHNTVMTALNRAVTSORLVKQNGRGRATRVLTANTWIRIDLRGQVHEVKKQMESSIEGKNTFPQCFRMYKSEIEQBEALNAADSRGLAKRFLRSGSGEHPVADYEAPAKISHGT"
CDS	complement(3588. .4250)
gene	/gene="upb"
CDS	/note="XAC3608"
gene	complement(3588. .4250)
CDS	/gene="upb"
gene	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
CDS	/codon_start=1
gene	/transl_table=11
CDS	/product="maleylacetate isomerase"
gene	/protein_id="AA038451.1"
CDS	/db_xref="GI:21109987"
gene	/translation="MSAALBELFSYWRSSAAVVRIGLQKALDYIAQVHLVRDGEQHAFAVQALNPOELVPTLHGHGVIAISALIEVEEAPADPAALIPRAFAEAAVRALAOIACDVPLNNLRVMQVLEBEFALDARQOQTRRMWEGFALLETOLARDQTRFCHGAPGLADCVLIPOLYNARREFVDLAPVTLRIIDQACIQLPAPFAAREMQADA VA"
CDS	complement(4282. .5277)
gene	/gene="upb"
CDS	/note="XAC3609"
gene	complement(4282. .5277)
CDS	/gene="upb"
gene	/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
CDS	/codon_start=1
gene	/transl_table=11
CDS	/product="fumarylacetate hydrolase"
gene	/protein_id="AA038452.1"
CDS	/db_xref="GI:21109988"
gene	/translation="VVMKLGSLKGGRDCLVVVARDIRTAVRATGIAPTLQRALEDWSNVAPRLNALSINLDSGADGVFDVQALAAPPRAYEFGVGSVYLPHVKEVRARNAEVPSFTYDPLMYQATGAGFPRDALKVISEDGIDLEAIIVVTDVPGASPEQAAAHQIOLGVINDVLSIKMLIPAEKAGGFGPQSKRSALSPFVTPDELDAMRSQKLHLPIYTHNGAFGQPEAGQNMOPQADLIHAATRTPLSGTIVSGCTIANODPTTGLASCPAQRQVETLRDKESTPFMAAGDVRIEMPDAASASIFGALQEVVERQPLP"
CDS	5654. .7036
gene	/gene="upb"
CDS	/note="XAC3610"

CDS 5654..7036  
 /gene="rhle"  
 /note="identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer/Genemark"  
 /codon\_start=1  
 /transl\_table=11  
 /product="ATP-dependent RNA helicase"  
 /protein\_id="AAM38453.1"  
 /db\_xref="GI:21109989"  
 /translation="MSFSLGLAPFLRLALAEQGYETPTPIQQQAIPLVLAGHDLLAG  
 AQTGKTAATGLPLLOHLGTAPVTNGPRXPRALILPTRELATOVHDSLRGYSKYL  
 RPSAVYGVGMGNQDLARRGVDLLIACPRGLIDHIERESVDLSGLVILDEADR  
 MLDWGLFSIKRILTKLPRQDRQILFSATFEENIKOLAEPMNPVOIQVTPNTVA  
 ESIITHRVPDGARKRDLLHLLAQDSREQLTVFARTKHGSKDLALFLERKGIKTAI  
 HNKSGQRMALSDFAKRVTLVATDIAARGIDIDQLPKVINYLPMVAEDVYHRI  
 GRTGRSTGSAISLVADAEKLFQIVRMILGRDVEIRDPVGPYPTPIRWGNSAPGR  
 AEOGCDRAPRKSHARRPHGDAPROAHAGPKPGQGRNGPRQANTGACAGRRDGG  
 RGGORPSPSTR"  
 complement(7238..9082)  
 /gene="XAC3611"  
 /complement(7238..9082)  
 /gene="XAC3611"  
 /note="identified by sequence similarity; putative; ORF  
 located using Blastx/Glimmer/Genemark"  
 /codon\_start=1  
 /transl\_table=11  
 /product="peptidase"  
 /protein\_id="AAM38454.1"  
 /db\_xref="GI:21109990"  
 /translation="MAATAMSPATWSTSTTALPKTKEMRATAMCARSSLRATRV  
 GAGSRGWRSMARSERSIPRATTALFRIPIMDRNWSVQSGVANFALYPGD  
 PLTPGVGASKAKRLIKIQAGSVLKIPILPSWDAQPLLAALGPVAPEDMRGALPI  
 TYRGDAAKARVHKVDADMGSTIYNIATLRGSEYDQWVVRGNHRDGVFGAADP  
 LSGTALLAEAKAGELAGOKPRKTLVYASWDGEAGLLSGTEWABOHADELRRKA  
 LVYNTDGGFLNAGSHALORLVNSVADVDODPSGVSVLERORARIRIDALAPA  
 AKPAKQDKAKTAAGGDIPLKALGSGSDYSPFLHGLATLIDLGVGGGSGVYHSL  
 YQSDYFARIDPQAYLPLLSQTVRLVLRNAPVLPQRFQFADNVAAGVAGELKQ  
 QADSDRTAARTQAEILLQAGVAAVDNPNRPORAPEKAAVPOIDFALDQAILQDS

Alignment Scores:  
 Pred. No.: 2e-05 Length: 12514  
 Score: 209.50 Matches: 110  
 Percent Similarity: 36.78% Conservative: 68  
 Best Local Similarity: 22.73% Mismatches: 164  
 Query Match: 10.23% Indels: 142  
 DB: 1 Gaps: 21

US-09-889-746-2 (1-400) x AE012011 (1-12514)

Qy 10 LeuSerLeuProLeuSerValAlaValThrGlnGlnLeuTyrAlaGlnProAlaGlu 29  
 Db 12317 CTGTCTCTTGGCGTGTGTGCTGTGATCCCG-----CTTGACATCGCCGACGAG 12267

Qy 30 SerLeuProThrValGluLeuGluProValValIleThrIleAspLysSerGlyMetAla 49  
 Db 12266 GCAACACCGCCGACC-----GGCAAGACCCCTGGCTACCGTCAGCGTCCACCGTCCCAAC 12213

Qy 50 LeuAlaAsnArgIleThrGlnMetProHisThrThrLysValIleTyrGluGluGlnIle 69  
 Db 12212 ATCAAGCGCAGCGACGACGAGCCCAACCCCGGTGTCAGTGATGACCGCCGACGACACTG 12153

Qy 70 GlnGluGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSer 89  
 Db 12152 GAGCAG-----TCCGGCAAGGTCAACCGTGGCCGACGCGCTCGGTTCATCTCGGCCAAC 12099

Qy 90 LeuGlyValSerSerGlyThrThrSerAnpHeGly----- 101  
 Db 12098 ACCGGCAACGCCCAATGAACCAACCAACCGCGTGGCGTCCGGCTCGCGGGGCATC 12039

Qy 102 -----GlnThrMetHisGlyArgGlnValGlnPheLeuLeuAsn----- 114  
 Db 12038 GGCCTGCGCGGACTGTCCGACAGAAACACGCTGGTCTCTCAACGGTCCGCGCTGGCC 11979

Qy 115 -----GlyValProLeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAsn 132  
 Db 11978 AATTACGGCTTTCCCGCOAATGGCTGGCGCACACCTTCGTCAATCTCAACGCGCTCGCG 11919

Qy 133 ProAsnGlnValAlaArgIleGluValLeu-----SerGlyAlaThrSerIleTyrGlySer 151  
 Db 11918 CTGGTGGCAGTGGAGCGCATCGAAGTGTCAAGATGGCGCTCGCGGTGTACGGCTCC 11859

Qy 152 GlyAlaThrGlyGlyLeuIleAsnIleValThrLysSerAspLeuGluGluGlnPhe 171  
 Db 11858 GACGGGTGGCTGGCGTCAATATCATCATCCCGCAGGATTTTCAGGGCGCCGAA--- 11802

Qy 172 GluThrArgIleGlyValHisGlySerLysLeuSerSerGluGlyIleGlyTyrGlnVal 191  
 Db 11801 -----CTGGGCTCAGCGCGGTACCTCCGATCAGGGCGGCATGGACACGCGCGT 11751

Qy 192 GlyGlnSerValAlaGlyVal-----SerGluAsnGlyAsnValLeu----- 205  
 Db 11750 GCCAATTCTGGCGGGATCGCGATCTGGACAGCGCGGTTCACATCATCTTTTCAGC 11691

Qy 206 -----AlaArgLeuAspValAspTyrArg----- 213  
 Db 11690 TTCGACGCTTACACCGCGACCGTCTCGATCAGGACCGCGCATCACCCTCGGGC 11631

Qy 214 -----ThrThrGlyGlyAlaPhe 219  
 Db 11630 ATCTATAGCATCTCCCGCGCGCTGGAATGCTGGTCCGCCAAGGGCGCATCTTAC 11571

Qy 220 AspAlaAsnGlyLysArgIleAlaProGluProAlaGlnThrAspLysGlnAspSerLys 239  
 Db 11570 CTGGTGGCGCGCTCGTCCGATGCTCGATCGCGAGCAACTGCCCGCGCGCAGC 11511

Qy 240 SerLeuSerValAsnThrAsnValAspTrpGlnLeuAspAspLysGlnAsnIleAsnLeu 259  
 Db 11510 GTGCTGACCGCGCGCGCATCGACGGGTGCGCGCGACACTGCGGGTTCACACG 11451

Qy 260 Ala-----LeuThrHisTyrAsnAspLysGlnAsp----- 269  
 Db 11450 GCGCGCTTACCACGCTTGATTCTCTACCAAGCGCTATAGGCGCTACACCAACGCC 11391

Qy 270 -----ThrAspTyrAlaProAspTyrGlyAsnArgLeu----- 280  
 Db 11390 TTCGGCTCAGCGATTCTGTGAGCGGTTCGGCGAGCGCTCTACAGGAGGTCAAGGCG 11331

Qy 281 AlaValLeuPheGlyGluLysPro-----SerLeuAsn 291  
 Db 11330 GTGTCGATCTTCCGCTCCAGCCCGTACTTCACTTGGAGGCGCGCTTTCGCTTGAC 11271

Qy 292 AlaIleLysGlyLeuSerLeuSerGluGlnProLysThr-----ThrLysSer 307  
 Db 11270 GCCAGCACTGTCTGGCC-----GAGCGGTCTCCAACTGTCTGCGCGCTTCCAGC 11220

Qy 308 ThrPheAsn-----IleAsnTyrHisHisAspAspLeuTrpGly 320  
 Db 11219 CCCTACACCCGTCAGCGCACCGCGGTGCGGATCGAATACACTTCTTTCGACCTG---GGC 11163

Qy 321 AsnThrIleAsnThrAsnAlaTyrTyrArgArgGluLysGlyArgPheTyr----- 337  
 Db 11162 CAGACGGTCAAAACCAAT-----CGCTCGAGTCTCTACCGCTTCTCGTGGCGGTG 11112

Qy 337 ----- 337  
 Db 11111 CGCGCACCAACGACGCTGGGACTGGGAGGCGCGGCTTTCGCGCACGACGACCCAG 11052

Qy 338 -----ProPheValAlaProPheSerIleAlaLysAlaLeuPro----- 350  
 Db 11051 CACGAAACCGTGTCCGCGGTTCGCCAATCGTTCACCTTGGCGCAGCGCTTGGCCGAC 10992

Qy 351 -----IleLeuGlnSerMetAsnLeuProSerAlaThrLeuAsp-AlaTyrTh 366  
 Db 10991 GGCAGTTACACCTGCTTGATCCAGCGCGACACCGAGCGGTGATCGATCGGCATCAAC 10932

Qy 366 rlyAlaPro 369

Db 10931 CTCGACCT 10922

RESULT 30 PGU87395 2050 bp DNA linear ECT 18-NOV-1999  
LOCUS PGU87395 complete cds.

DEFINITION Porphyromonas gingivalis TonB-dependent receptor Hmnr (hmnr) gene,  
complete cds.

ACCESSION U87395  
VERSION U87395.2 GI:6449463

KEYWORDS

SOURCE Porphyromonas gingivalis.  
ORGANISM Porphyromonas gingivalis.  
Bacteria; Bacteroidetes; Bacteroidales; Porphyromonadaceae; Porphyromonas.

REFERENCE 1 (bases 1 to 2050)  
Simpson, W., Wang, C.Y., Mikolajczyk-Pawlinska, J., Potempa, J.,  
Travis, J., Bond, V.C. and Genco, C.A.  
Transposition of the endogenous insertion sequence element IS1126  
modulates gingipain expression in Porphyromonas gingivalis  
infect. Immun. 67 (10), 5012-5020 (1999)

TITLE

JOURNAL MEDLINE 99426790

REFERENCE 2 (bases 1 to 2050)  
Simpson, W., Wang, C.Y., Forng, R.Y. and Genco, C.A.  
Direct Submision

AUTHORS Submitted (28-JAN-1997) Microbiology/Immunology, Morehouse School  
of Medicine, 720 Westview Drive, Atlanta, GA 30310, USA

JOURNAL 3 (bases 1 to 2050)  
Genco, C.A., Karunakaran, T. and Simpson, W.  
Direct Submision

REFERENCE 3 (bases 1 to 2050)  
Submitted (18-NOV-1999) Medicine, Boston University, 774 Albany  
Street, Boston, MA 02118, USA

AUTHORS Sequence update by submitter  
JOURNAL Location/Qualifiers

REMARK 1. 2050  
COMMENT /organism="Porphyromonas gingivalis"  
FEATURES /strain="A7436"  
/db\_xref="taxon:837"

110..2050  
/gene="hmnr"

110..2050  
/gene="hmnr"

110..2050  
/gene="hmnr"

110..2050  
/gene="hmnr"

110..2050  
/gene="hmnr"

110..2050  
/gene="hmnr"

110..2050  
/gene="hmnr"

110..2050  
/gene="hmnr"

110..2050  
/gene="hmnr"

110..2050  
/gene="hmnr"

110..2050  
/gene="hmnr"

110..2050  
/gene="hmnr"

110..2050  
/gene="hmnr"

110..2050  
/gene="hmnr"

110..2050  
/gene="hmnr"

110..2050  
/gene="hmnr"

110..2050  
/gene="hmnr"

110..2050  
/gene="hmnr"

110..2050  
/gene="hmnr"

110..2050  
/gene="hmnr"

110..2050  
/gene="hmnr"

Qy 2 ArgHisSerHisIeTyPheGlnTrpLeuSerLeuPro-----Val 13

Db 53 AAAAGAGATCATCTTACAC---TTATACATACCCCGTTAATATAGTAGAGGAAT 109

Qy 14 LeuLeuSerValAlaValThrGlnGlnLeuTy----- 24

Db 110 ATGAAAGGTAGTAAACAAAGAGAGCCCTCATCGGCTCTTTCTTATATATAGTATA 169

Qy 25 -----AlaGlnProAangIuSerLeuProThr-----Val 34

Db 170 TACTCCATCGGACCAACCTCCGCGCACTACCGACCATGATCCGCAATATC 229

Qy 35 GluLeuGluProValIleThrIleAspIysSerGlyMetAlaLeuAlaAsnArgIle 54

Db 230 GCACCTGAGATATATAGTGTAC-----GGTAGCCGATACAGCCGCTGCTT 277

Qy 55 ThrGlnMetProHisThrThrIysValIleTyrgluGluGlnIleGlnGluGlnAlaThr 74

Db 278 AAAGATGATCTGTCGCCCAAGAGGTTCAGAGCCCAAGATATCAAGCTTATAGCCCA 337

Qy 75 GlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSerSer 94

Db 338 TCTTCT-----TTCAATGACGATGACGATATATCTTCCGCGATGAGTT----- 385

Qy 95 GlyThrThrSerAspPheGlyGlnThrMetHisGlyArgGlnValGln----- 110

Db 386 -----ACCAAGCATGTGTCAGAGATCAGTCAATGCTCAG 421

Qy 111 -----PheLeuLeuAsnGlyValProLeuThrGlySerArg 122

Db 422 GATTTAGCAAAAGTTCTATCTCTCTCTCGTATGCGAATGATTTCAACGGATCT 481

Qy 123 AspIleSerArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGluValLeu 142

Db 482 ACCAGTGAAATGACTTCAACGAAATCAATCCGATGACATGAGGAAATCAAGTCTT 541

Qy 143 SerGlyAlaThrSer---IleTyrglySerGlyAlaThrGlyGlyLeuIleAsnIleVal 161

Db 542 CGGAGAGCTTCTCTCTCTTGTACGATATATGCAATGAGAGGTATATCAATATCATC 601

Qy 162 ThrIysSerAspLeuGluGluGluGlnPheGluThrArgIleGlyValHisGlySerIys 181

Db 602 ACCCGTACA-----GCCAAGATCCTTT-----CGGATATCTGCTTCGCTCGATAC 649

Qy 182 LeuSerSerGluGlyIleGlyTyrglnValGlyGlnSerValAlaGlyValSerGluAsn 201

Db 650 GATATGTCGCGACGACGAAATACGATGTG-----GCAGCAGAGAGTG---AAACGT 697

Qy 202 GlyAsnValLeuAlaArgLeuAspValAspTyrglnThrThrGlyGlyAlaPheAspAla 221

Db 698 GGGATTTTACCATGATGATGCGCGGATCAATATAGCAGCAAAAGTTATATTTGGCC 757

Qy 222 AsnGlyIysArgIleAlaProGluProAlaGlnThrAspIysGlnAspSerIysSerLeu 241

Db 758 GAT-----CAGTTCCAGCAGCAACTG 778

Qy 242 SerValAsnThrAsnValAspTrpGlnLeuAspAspIys-----Gln 255

Db 779 AAGGTTCAGGCAATACGATGATATATATACAGAAATTTACATCTCTCCGACGAA 838

Qy 256 AsnIleAsnLeuAlaLeuThr-----HisTyrglnAspIys 267

Db 839 AACTTATCTTCACTTACCTTACCGCATGTCGATCTGCGTACGACGATTAAG 898

Qy 268 GlnAspThrAspTyrlaProAspTyrglyAsnArgLeuAlaValLeuPheGlyGluIys 287

Db 899 ATCGATTTTCTGATC-----AACTCC 919

Qy 288 ProSerLeuAsnAlaIleIysGlyLeuSerLeuSerGluGlnProIythrThrIysSer 307

Db 920 TACGATGCAAAAGCCGTCGCAACGTCGATCAGCAG-----ACTTCG 964

BASE COUNT 568 a 501 c 445 g 536 t

ALIGNMENT SCORES:

Pred. No.: 2.24e-06 Length: 2050  
Score: 209.00 Matches: 91  
Percent Similarity: 39.30% Conservative: 114  
Best Local Similarity: 24.66% Mismatches: 110  
Query Match: 10.21% Indels: 18  
DB: 1 Gaps: 18

US-09-889-746-2 (1-400) X PGU87395 (1-2050)



```

/translation="WPKKKQVINTKNAIPALGPSQAIENGVLTYSQIDIPFVATMT
LVSEDEVEQTKQSLNIGAILKEAGYDPRDVVASATVYIKMDNDFKINGVDKYLGEV
KPARACVEVARPKDVKEIGVIAVK"
complement (7314..10142)
/gene="FN1974"
complement (7314..10142)
/codon_start=1
/translate=11
/product="DNA/RNA helicase (DEAD/DEAH BOX family)"
/protein_id="AAL94064.1"
/db_xref="GI:1971231"

```

```

/translation="WSTIDLEALRTSSIDPNIDSEKQYELIANGSEKIVTRIKYF
EDDEFISVARTHTWGISLFLBLKMNENKIKK1ITGDLTFTEPKALKLSYK
NIDLKVTNRKHTWAVFPRKGNWTLTVGSENLTQGLTVFENNIVNSENKIV
KSVLEFNRFPNLKTLIEDIENTOKYEOLEKLIENNNOLDLDEIKPNMOVAL
KULERTRENNBALIISATGCKTYSAPDYQAKAKLIPAHKHLSEKISYOR
ILAKKMEI FDNKPOINDKDEVPMVOTLKEKLNLPKDPYPIYIIDEVHGAK
TYOSIFEFKPEFLGITATPERDTPNLYQFNVAVERLQDAMEBELCPHYF
GISDIVDGESEIDERTSLAKLSDTRVKHILSENSRYASISGRSLCLFVSKVEAKI
LVKFLQGVKAIALSSENSDNREEARIKLESGEIEYTSVIDFNEGVDLPCVQVI
LIRPTSAIVYIQGLRGRKHKNDYTVLDPFGNENKFLIPAISSONSYKDM
KRELMADPFLAGSESISEDESKERIEENIKTPSRKLIIEPKLLEKQIGIPY
LYDPERKMLSPYLVILKTKKIDVYLKINAPRYKANNIEKFLIFUSTFPTPAK
IHEMLIKELIKOLNLIETVKLIKDKKSLNNQNSIRNAEHLSEKIFITLSTTKS
FDPVLYKDKKEYLIDENPNKSNYSYFELIDDLIKYNAFAERNYNNFVESYKLF
GSYTKQBAFWYLNPNPNFQVSGYTPENREKLIIFITMDLSRADYNSFYSOT
FSMFSSKSYLKKDKLITEGLIENPFEINVFVKNGENNYIYGVDEKVLASEIK
DSQGSMTKYLIFKLDKDYKELLDYFNN"

```

BASE COUNT 3309 a 1750 c 980 g 4194 t  
ORIGIN

## Alignment Scores:

Pred. No.:	3.64e-05	Length:	10233
Score:	204.50	Matches:	93
Percent Similarity:	41.588	Conservative:	60
Best Local Similarity:	25.274	Mismatches:	98
Query Match:	9.994	Indels:	117
DB:	1	Gaps:	21

US-09-889-746-2 (1-400) x AEO10499 (1-10233)

```

QY 14 LeuLeuSerValAlaValThrgIngluLeuTyraIgluProangluSerLeuProthr 33
DB 5704 TTATTGCTTAATGTATATAGAGAGAGATGCTTTTGACAGAAAGT---CTGTT 5648
QY 34 ValGluLeuGluProValValIleThrIleAspLysSerGlyMetAlaLeuAlaSerArg 53
DB 5647 TAGAGCTTAACAAACAGTTGTAACCTTCGATGCTTTGGG-----5606
QY 54 IleThrgLmePProIleThrThrlvs-----ValIleTyrgLugluInlIegin 70
DB 5605 ---ACACCTGTTGTAAACAGCAAAATAATATACAGTTRITTAATGCAAAAGAAATAA 5549
QY 71 GluGluAlaThrgLysSerArgInleuAlaAspValMetAlaGluLeuIleProSerLeu 90
DB 5548 GAAAAA-----GTCGCAAAACTATTGCTGATGACCTT---AGAGAGATACAGGTT 5498
QY 91 GlyVal-----SerSerGlyThrThr 97
DB 5497 GTTGTAGACAAATGATGAGCTTCGCTATGTAGATTAAAGGGGTTACAGGCAACA 5438
QY 98 SerAsnPhagLysInlThrMetHisGlyArgGluValGlnPheLeuLeuAnglyValPro 117
DB 5437 TCACATTTAATCTGTATT-----TTACTAGATGGATATCCA 5399
QY 118 LeuThrgLysSerArgAspIleSerArgGluLeuAsnSerIleAsnProangluAla 137
DB 5398 GTAAAGTGAAGCTTCAGGCTTAAC-----TTAAATACCTGTTCCAGTATGAATGA 5345
QY 138 ArgIleGluValLeuSerGlyAlaThrSerIle---TyrgLysSerGlyAlaThrGlyGly 156
DB 5344 AATAATGAAGTCTTCAAGAGAGAGAGATTATGTATGTAGATGAGTATGTTGGG 5285

```

```

QY 157 LeuIleAsnIleValThrlvsSerAspLeuGluGluInlPheGluThrArgIleGly 176
DB 5284 GTAGAAATATATATACAAAGCA-----CCAACTAAAAAAGCT 5246
QY 177 ValHisGlySerLysLeuSerSerGluGlyIleGlyTyrgInlValGly----- 192
DB 5245 GTCATATGTT-----GGAGCAGATTAAGAGTGGTTCCTTGAGAACT 5204
QY 193 -----GlnSerValAlaGlyValSer 199
DB 5203 ATTAAGAGAAATGTTTATCTTGCTGGTAAAAATTGAGATTAATTTCTGTAAATGCTTCA 5144
QY 200 GluAnglyAsnValLeuAlaArgLeuAspValAspTyraArgThrThrgLysAlaPhe 219
DB 5143 TATCTCGAAAT-----TCAAGTAAAGATTATTAAGATGAAGAAGTCCAAATAT 5096
QY 220 AspAlaAsnGlyLysArgIleAlaProGluProAlaGlnThrAspLysGluAspSerLys 239
DB 5095 GAA---AATTAAGAA-----GATMAAAGAGATAGCCTT 5066
QY 240 SerLeuSerValAsnThrAsnValAspTrpGluLeuAspAspLysGluAsnIleAsnLeu 259
DB 5065 TCGTTAAGAGAGAAATATTATTAGAT-----AATGGAAGTATA 5027
QY 260 AlaLeuThrHisTyraAsnAspLysGlnAspThrAspTyraAlaProAspTyrgLysAsnArg 279
DB 5026 GCAATT---AATTATATCATGTAGATGAATGAATAT-----4991
QY 280 LeuAlaValLeuPheGlyGluLysProSerLeuAlaAlaIleGlyLysLeuSerLeuSer 299
DB 4990 -----TATACAGGCTCTTATTAAGAGAGAAATTTGAT 4958
QY 300 GluGluProLysThrThrlvsSer-----ThrPheAsnIleAsnThrlHis 315
DB 4957 AAAAAATCCAAAGCAAGTAGGCTCTTGAGCGGTTACCTATGATTAATAAGACTTATTC 4898
QY 316 AspAspLeuTyrgLysAsnThrlleAsn-----ThrasnAlaTyrg 328
DB 4897 AATGCAAAATATATATCAAAATAATATGATAAATGATATCTTCTTGACAAAGTGATAT 4838
QY 329 TyraArgArgGluLysGlyArgPhe 336
DB 4837 TAT---CATATATAAAATAATATTC 4817

```

RESULT 32

AF048749/c 24454 bp DNA linear BCT 23-JUN-1999

LOCUS Bacteroides fragilis capsular polysaccharide C biosynthesis operon, complete sequence.

ACCESSION AF048749

VERSION AF048749.1 GI:5199106

KEYWORDS Bacteroides fragilis.

SOURCE Bacteroides fragilis.

ORGANISM Bacteroides; Bacteroides; Bacteroidales; Bacteroidaceae; Bacteroides.

REFERENCE 1 (sites)

AUTHORS Comstock, L.E., Coyne, M.J., Tzianabos, A.O., Pantosli, A., Onderdonk, A.B. and Kasper, D.L.

TITLE Analysis of a capsular polysaccharide biosynthesis locus of Bacteroides fragilis

JOURNAL Infect. Immun. 67 (7), 3525-3532 (1999)

MEDLINE 99307214

PUBMED 10377135

REFERENCE 2 (bases 1 to 24454)

AUTHORS Comstock, L.E. and Coyne, M.J.

TITLE Direct Submision

JOURNAL Submitted (17-FEB-1998) Channing Laboratory, Brigham & Women's Hospital and Harvard Medical School, 181 Longwood Avenue, Boston, MA 02115-5804, USA

FEATURES

source 1..24454

```
/organism="Bacteroides fragilis"
/strain="NCTC 9343"
/db_xref="taxon:817"
complement(<1..966)
/note="orf1"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAD40705.1"
/db_xref="GI:5199107"
/translation="MNNVQOQVTSORIKSIDPLFLMAGGAALLSYSLVYALRKPYYTA
AGFQGLEAFGDKYKVYVYIAQILGYLSKPIGKILSELKRENMKFLISILAEAS
LILFELLPAFNTGAMFLNGLSGCMGGLIFSFIEGRMTDILASLGVSNVISGTA
KSAGLYVNDTINSEFMFNPALIGVALPALLGYALNRLPOPTAEDIAMKSKRETLN
GKQWLEFKNMFPFLTLFIANVLTILRDKEDFLVKIIDVQYSSMFAQVDSVVT
LILILFGLMVFVRSNLKALSILLGLIIASVMVMAVVSFGYEQLQL"
1272..1907
/note="orf2"
/codon_start=1
/transl_table=11
/product="putative methyl transferase"
/protein_id="AAD40706.1"
/db_xref="GI:5199108"
/translation="MNKRGFVSRILONFRKPEGFGRMLAGMTGHASLAQMGSCIL
QQPWESVIDIGCGGANLLQLQCPQGAHQADIDISPESTVPARKONKKYLGRCPFI
EQGGVHRFLPYDPAFDAYFATFVYFWNGLOHAFTEVARVLKPGGSFLICCEISDPAN
KAWTGLVGMIEIHSDELKAILSKSGFTDTAIFRTKKEELCLVSHRQTURL"
2798..3139
/genes="upcy"
/2798..3139
/genes="upcy"
/note="formerly orf3"
/codon_start=1
/transl_table=11
/product="upcy"
/protein_id="AAD40707.1"
/db_xref="GI:5199109"
/translation="MVFVHISRQEMAARVSRNLNVDFLRDRSTGAPTCPIDQAWADFR
YMLDSQOQVILTGESSIPKTRVVVAKGDLQGLRGELVRYNNKYHILVRIDMFGSAM
TPASVYRKEK"
3322..3714
/genes="upcz"
/3322..3714
/genes="upcz"
/note="formerly orf4"
/codon_start=1
/transl_table=11
/product="upcz"
/protein_id="AAD40708.1"
/db_xref="GI:5199110"
/translation="MKNLSICSVSSVVKLLLEYGBDGTVPVYVNELTALNQELRLCAD
LILQKGEPEBEAELVTLFKGYDTMLFNFSSENEQVIQELLDRSMTVLEKLPASVLK
COLLECFQOTGEELIREAKKNIERVI"
<3717..19221
/genes="capsular polysaccharide C biosynthesis operon"
/note="transcriptional start has not been mapped so first
gene in operon is unknown"
3717..4604
/genes="rmlA"
/3717..4604
/genes="rmlA"
/codon_start=1
/transl_table=11
/product="glucose-1-phosphate thymidyl transferase"
/protein_id="AAD40709.1"
/db_xref="GI:5199111"
/translation="MKGIVLAGSGSTRLYPITKGVSKQLLPIFPDKPMIYYPISVMLLA
GIRELITSTPDPLFGQRLQDGSDFGVRFEYAEQSPDGLAQAFIIGEXFNGDVS
CIVLGDNTFYGGQFTRLMNEAVRIAESSEKATVPGYVSDPERYGAVFEDNGNVSFI
EKPKQKSNYAVAVGLYFYPNKRVVAKSIRPSRGELEITTVNQNFSLDKELRYQLL
GRGFALWDGTGTHDSISEASTFIEVIEKRGQGLKAVACLEGIALRKKGWISPEKMKALAOQM
LKNQYQVLLKVIDELVYK"
4620..5168
/genes="rmlC"
4620..5168
/genes="rmlC"
/codon_start=1
/transl_table=11
/product="dTPP-6-deoxy-D-glucose-3,5 epimerase"
/protein_id="AAD40710.1"
/db_xref="GI:5199112"
/translation="MNIITAEGLVLEPQLFKDDRGVYFFESFNOREPEEKVKTTFF
VQNESKSSYGVIRGLHFQPPPAQKLVRIKGAVLDAVDIRKGSPTFGKHVSVEL
TDENHQFFTPRGFAHGFSLVSEVIFQVKCDNFYHPEABEAIAMNDPDLGIEWRVPC
KSILSKDSRHPLLKXVILFK"
5187..5708
/genes="wcfA"
5187..5708
/genes="wcfA"
/codon_start=1
/transl_table=11
/product="putative acetyl transferase"
/protein_id="AAD40711.1"
/db_xref="GI:5199113"
/translation="MKKLVYDVRKDIIPMNSRITRILFPLNKIPYLRLFSKPLLRKQF
NLGNTVQFNSGFFCHAPKLCGNVGLSDTFILAYADVIGNNVSFSFRNMLITSTHD
VNDENKTIASSIVIGNNVWITSVILLAGVKIGDNTIIGAGSVVTHDIPANVPAAGNP
CRVKSIRFNKNE"
5701..7236
/genes="wzx"
5701..7236
/genes="wzx"
/codon_start=1
/transl_table=11
/product="putative flippase"
/protein_id="AAD40712.1"
/db_xref="GI:5199114"
/translation="MNSRTAKGLKNAQVALLFVFINLILSLFSRKAFIEHLGAELVG
LNTTLYLNLAEGLGVAISYALYKLYEGDKQAVNELISIQGLMKRIAIIVI
IMSGVILLFPFPFAKIEIPWYATFLVLSSSFYSYFNKIQFVLTAQKEYKLI
TNQGVKVTILQIVVYFCNGYVWMLLELMGVITFVLNSIVRKEYPMLQTSIP
KIGLLNSVFSYIIRKTKQLFHFKIANVLNQTSPIIYSYTNLTWVAVGVNMYLIIS
GIDLLNSVFSIGAGIGNLVABGNQAKIQVFNELLSIWTISVILCFGVYQMSRPF
ILKMWGRFVLDLDFYLLMLLIAFISLTRLVDLPAAAGLYQDVMAALBAFLNLGLS
ILKVGWGLSGILGGVTSIVIIALLMKPFELPKYGNKCLAFYFVYMKCVAPALIT
FYRSIRVDYIGIGMCTDYSWILLISMLNIPVTIISPPIFFPSDGTNRFKRVINI
VFN"
7240..8103
/genes="wcfB"
7240..8103
/genes="wcfB"
/codon_start=1
/transl_table=11
/product="putative fucosyl transferase"
/protein_id="AAD40713.1"
/db_xref="GI:5199115"
/translation="MLVILRGLRGNLFOIATAASLTQNFICTVKNKDOEROVLLYK
DSFFKNVKNMGVDPDGIYYKEPFHESRIPYEGKDLIIDGIFQSKYKRSVSLDL
YRITDEURKKIWNICGNILEKGETVSIHVRG DYKLPHALPFCGKSYKNAIQIYGE
DKPIICSDIDIDCKCKNFIGKRYFIENFTPLDLYTQSLCTHNIISNSFSWAGWL
NENSKIVIAPOWFWGISVKLGVSDDLPSVRLPNNTYLTGRYCFALYKVVEDYLLNI
LRLIWKBKKNM"
8079..9239
/genes="wcfC"
8079..9239
/genes="wcfC"
/codon_start=1
/transl_table=11
/product="putative glycosyl transferase"
/protein_id="AAD40714.1"
/db_xref="GI:5199116"
/translation="MEKKEYKXKKMTLIVNHPMNDVFEGKDVFLVFCYIGKIYHYDV
KIVFPYSYLAEAKTIRDELVELPVSYIKKLAPFSFVFGKLMRYLFCNARKTDLMPRH
VTITATMIITIKMLNENGIAIKLDSGNVNFYKNNLSKSLFKLLYRRMFLVDVF
SYETKMGLENTQOTIGIDISSKLFYMPNGFDEDMIKHLDINVKESYDKENWMTVGR
LGTNQKTEFLKXAVENIDLKDWVIYILIGDIDPQNTIFLEPVSNFFSNHPEKKSVIP
```

```

gene
CDS
    TCATSDKRHLMEYTKSKVFLVTSRCESGLVLENAKRFNFTVSTNCAFEEDLVESG
    KYGEI.PQDNTITLACILEKIIILGOLDIDVNDPSPESLSTYYQVKKMKNN"
    9257. .9913
    /gene="wcfD"
    9257. .9913
    /gene="wcfD"
    /codon_start=1
    /transl_table=11
    /product="putative acetyl transferase"
    /protein_id="AAD40715.1"
    /db_xref="GI:5199117"
    /translation="MHTYLPRLCYCKRFLVFPVCTSKTENTNRYFKRIKPLP
    SRIVTFEGHVDKSIIVPEYIRGSKIQLQINFRALSGLEALIDFNSSHPSTI
    VIGDVFNNSCHIGATNRIIGNDVVAARVFIIDHFRGTTVDLQIPVNRLLYS
    KGPVVIIGNVWIGEGASILPNVTIGDINSIVAAHVATKNI PANSIAGVCPARIIRTIN
    "
    9914. .11011
    /gene="wzy"
    9914. .11011
    /gene="wzy"
    /codon_start=1
    /transl_table=11
    /product="putative polymerase"
    /protein_id="AAD40716.1"
    /db_xref="GI:5199118"
    /translation="MNLFLVLFVFGIIGSYSKSEKRLFWLEFVIAIITGRSI
    SVPDIOYQIOLSSILIDIPLYSVEIGOLTKIKLTNNSSPFYFGIILVINSYF
    LKFSFCPTDQSPYKSSGTLINQIKQIEFLILVPSVGLVYTGTIVLRAGIAMSLIL

```

```

Alignment Scores:
Pred. No.: 0.000153 Length: 24454
Score: 202.50 Matches: 89
Percent Similarity: 40.15% Conservative: 72
Best Local Similarity: 22.15% Mismatches: 153
Query Match: 9.89% Indels: 87
DB: 1 Gaps: 17

```

US-09-889-746-2 (1-400) x AF048749 (1-24454)

```

QY 11 SerLeuProLeuSerValAlaValThrGlnGlnLeuTyrAlaGlnProGlnSer 30
DB 23407 ACCTTCCTTTACACAGAAATGTCGGAAGAACAGACACAGCTCTTATATGAC 23348
QY 31 LeuProThrValGlnLeuGlnProValValIleThrIleAspLysSerGlyMetAlaLeu 50
DB 23347 GATGATCCTTCAATCTTGAACAATTTGTGTTACGCGAACCCGACAA----- 23300
QY 51 AlaAsnArgIleThrGlnMetProHisThrThrLysValIleTyrGlnGlnIleGln 70
DB 23299 GAAAGAGAGATTAAGAACACACCGGTCATCATCATATATCATCTTAAGCAATAGAA 23240
QY 71 GlnGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeu 90
DB 23239 GAAAGAGAGACCGGTACATCTAG-----GACCTTCTGATCAAGAGCTCCCGACATT 23186
QY 91 Gly-----ValSerSerGlyThrThrSerAsnPhgGlyGlnThrMetHisGlyArg 107
DB 23185 AACTTTCAGAGGTTGGCTATGAAACGACATGATATA---CAGGATTAAGTTCCAAA 23129
QY 108 GlnValGlnPheLeuLeuAsnGlyValProLeuThrGlySerArgAspIleSerArgGln 127
DB 23128 CATATCCTTTCTGATATAGACGGCGAAGCATATAGCGGCAAAAAGGTGCAACATCGAC 23069
QY 128 LeuAsnSerIleAsnProAsnGlnValAlaArgIleGlnValLeuSerGlyAlaThrSer 147
DB 23068 TATTCGGAGATCAATCTTTATATATATGACCATATCAAAATAGTCAAGAGCTTCTTCG 23009
QY 148 ---IleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLysSerAspLeu 166
DB 23008 GCCCTCATGTTCTCAAGGATGGGAGGATTCATCAATCATATACCGCTAAA----- 22955
QY 167 GlnGlnGlnGlnPheGlnThrArgIleGlyValHisGlySerLysLeuSerSerGlyGly 186

```

```

DB 22954 GCCAAAAAGAAATTCGAGGCTTCGCGAGCATACGCTATGACGAGAAACAGCAAAAC 22895
QY 187 -----IleGlyTyrGlnValGlyGlnSerValAla 196
DB 22894 TATTAAGATACCTCCCAAGATCATTCGCAATACAAATATGCAATTCATCGATTAACCC 22835
QY 197 GlyValSerGlnAsn-----GlyAsnValLeuAlaArgLeuAsp 209
DB 22834 AATCTGAACACCAATCTGCTCTTGATTTGAACCTGGCAAGTTCCACATGAACCCGAC 22775
QY 210 ValAspTyrArgThrThrGlyGly-----Ala 218
DB 22774 GACCTTTACAAAAGTTTCGATGATATACCAATATTCGATTAACAACTCTCGTGAATAAT 22715
QY 219 PheAspAlaAsnGlyLysArgIleAlaProGlnProAlaGlnThrAspLysGlnAspSer 238
DB 22714 TTTCCGCTTATTAACCAACCAATTTACCGAAGATCTCAGTAAACCCGACGATATATCG 22655
QY 239 -----LysSerLeuSerValAsnThrAsnValAspTyrGlnLeuAspLysGlnAsn 256
DB 22654 GGATACGAAGAGGTACCAAGTACGCCATTAATGACATATCGTTTCAGCAACGGCTCMAA 22595
QY 257 IleAsnLeuAlaLeuThrHisTyrAsn---AspLysGlnAsp----- 269
DB 22594 GTCCAGCTTAAAGAAAGCATATATATGCTGCAACAAATATGATTTTCACACAGATATATA 22535
QY 270 -----ThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeuPheGly 285
DB 22534 TTCGAGAAATCAGAGGACATATACC-----TATGGCGAAGCATATGATTAACAGATTCC 22481
QY 286 GlnLysProSerLeuAsnAlaIle----- 293
DB 22480 GACAAATCTCTTTGTGTACCTCTGTTTCATACGATCATCAACCGATATGATTAATAC 22421
QY 294 -----LysGlyLeuSerLeuSerGlu-----GlnProLysThrThr 305
DB 22420 GAATCGAAGAGGCTGCTGCTCGATATTAACAAATATTTATCCAGCCCGCTCGTA 22361
QY 306 LysSerThrPhe-----AsnIleAsnTyrHisAsp 316
DB 22360 TATAGCATACCGCGCTCATTAACACAGACCATTAACGGGAGATGGAATTTACAGAGAA 22301
QY 317 AspLeuTyrGlyAsnThrIleAsnThrAsnAlaTyrTyrArgArgGlyLysGlyArgPhe 336
DB 22300 TCATTTTTCAGATATATTTGAAACCGGTGCG-----AAGAAACAAAGAACCATGG 22247
QY 337 Tyr 337
DB 22246 TAT 22244

RESULT 33
AL627276/c 274050 bp DNA linear BCT 06-JUN-2002
LOCUS Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
DEFINITION complete chromosome, segment 12/20.
ACCESSION AL627276 AL513382
VERSION AL627276.1 GI:16503805
KEYWORDS
SOURCE
ORGANISM
Salmonella enterica subsp. enterica serovar Typhi.
Salmonella enterica subsp. enterica serovar Typhi
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.
REFERENCE
1 (bases 1 to 274050)
Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
Sebaiha,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
Comerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
Farrar,J., Feltham,T., Hamlin,N., Hughes,A., Kien,T.T., Holroyd,S.,
Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moutre,S., O'Gaora,P.,
Parry,C., Quail,M., Rutherford,K., Skellern,M., Skellern,J.,
Stevens,K., Whitehead,S. and Barrett,B.G.
Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18

```







```

source
1. .2495
/organism="Escherichia coli"
/strain="CP9"
/db_xref="taxon:562"
/notes="human clinical isolate"
repeat_region
1. .137
/rpt_family="insertion sequence 1230B"
/rpt_type=dispersed
203. .2495
/gene="iron"
protein_bind
203. .221
/gene="iron"
/notes="putative"
/bound_moiety="Fur"
321. .2495
/gene="iron"
/codon_start=1
/transl_table=11
/product="siderophore receptor Iron"
/protein_id="AAF00587.1"
/db_xref="GI:6006729"
/transl_table="MRINKLWSLVLLVGLNSQVSKYSDDDNDDETLVVEATAEQV
LKQPGSVVITSEDIKTPPVNDLSDIIRKMPGVNLGNSASGTRGNRROIIRGMGP
ENTLIDGVPVTSRNSVYRWSRGERTRGDTWVPPQVERIEVIRGPAARVCSGA
AGVVNIITKRPITNDWHSUSLYTNQPSSEEGATRRANFSLSPLAGDALTTRLYGN
LNKTDADSDINSVPVGTNAAGHEGVNRKQDINGVSKLNPQQIQLDFEVGSRQNIY
AGDTQNSSAAVTESLAKSGKETNRLYRQNYGIHNGIWMQGSFRFYVEKTNTRM
NEGSGGEGRI LAGKFTNRLSSWRTSGELNPLNVMVDOTLTVGAENWRDKLDDP
SSTSLVNDRLDISGSAADRSKNHSOISALYIEDNIEPVCTNII PGLRPDLYSD
SGNFSPLNLSQBLGDFYKAGVARTFKAPNLYQSEGVLLYKNGNCPKDIITSGG
CYLNGKDLDEPISVNEIGLETFWEDTHASVYFRNDYQNKIVAGDNIQOTASGAY
ILKWQNGKALVDGIEASMSPLVKERLNMTNATMITSEQDKTGNPLSVIPKTYTN
NSLNTWITQAFSAFNFTLYGKQPRTHAETRSDDTGLSGKELGAYSLVGTNPNYDI
MKNLKLVGVSINLKNKIFRSSEGANVNEPRAYYAGVTASF"
BASE COUNT 686 a 557 c 632 g 620 t
ORIGIN

Alignment Scores:
Pred. No.: 1.44e-05 Length: 2495
Score: 199.50 Matches: 84
Percent Similarity: 41.94% Conservative: 59
Best Local Similarity: 24.63% Mismatches: 135
Query Match: 9.75% Indels: 63
DB: 1 Gaps: 16

US-09-889-746-2 (1-400) x AFI35597 (1-2495)

Qy 9 TrpLeuSerLeuProLeuLeuSerValAlaValThrGlnGlnLeu-----Tyr 24
Db 342 TGG---TCGCTAACTGTGCTCTAGTTGGTTGTAATAGCCAGGATCATGACCCAAATAC 398
Qy 25 AlaGlnProAsnGluSerLeuProThrValGluLeuGluProValValIleThrIleAsp 44
Db 399 TCCGACGATGATAATGAC-----GAGACTCTGTGTGGTGAA----- 434
Qy 45 LysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrThrLysValIle 64
Db 435 -----GCCACCGCTGAGCAGGATATTAAACACAGCAGCCGGCGTGTGCTGTTATT 482
Qy 65 TyrGluGluGlnIleGlnGlnAlaThrGlySerArgGlnLeuAlaAspValMetAla 84
Db 483 ACCACGAGGATATT---AAAAGACCCCTCCGGTAAACGACCTTTCAGATATTATTTCGT 539
Qy 85 GlnLeuIleProSerLeuGlyVal-----SerSerGlyThrThrSerAsnPhe 100
Db 540 AAAATG---CCTGGTGTAACTTACCGCAATAGCGCTCGGGCACACCGCGGTAAATAC 596
Qy 101 GlyGlnThr-----MethIleGlyArgGlnValGlnPheLeuLeuAsnGlyVal 116
Db 597 CGCCAGATCGATATTTCGTGGTATGGGGCGGAAACACCTTAATTTAATTGATGCTGTA 656
Qy 117 ProLeuThr-----GlySerArgAspIleSerArg 126

```

1. .2495  
/organism="Escherichia coli"  
/strain="CP9"  
/db\_xref="taxon:562"  
/notes="human clinical isolate"

repeat\_region  
1. .137  
/rpt\_family="insertion sequence 1230B"  
/rpt\_type=dispersed  
203. .2495  
/gene="iron"

protein\_bind  
203. .221  
/gene="iron"  
/notes="putative"  
/bound\_moiety="Fur"  
321. .2495  
/gene="iron"  
/codon\_start=1  
/transl\_table=11  
/product="siderophore receptor Iron"  
/protein\_id="AAF00587.1"  
/db\_xref="GI:6006729"

CDS  
/transl\_table="MRINKLWSLVLLVGLNSQVSKYSDDDNDDETLVVEATAEQV  
LKQPGSVVITSEDIKTPPVNDLSDIIRKMPGVNLGNSASGTRGNRROIIRGMGP  
ENTLIDGVPVTSRNSVYRWSRGERTRGDTWVPPQVERIEVIRGPAARVCSGA  
AGVVNIITKRPITNDWHSUSLYTNQPSSEEGATRRANFSLSPLAGDALTTRLYGN  
LNKTDADSDINSVPVGTNAAGHEGVNRKQDINGVSKLNPQQIQLDFEVGSRQNIY  
AGDTQNSSAAVTESLAKSGKETNRLYRQNYGIHNGIWMQGSFRFYVEKTNTRM  
NEGSGGEGRI LAGKFTNRLSSWRTSGELNPLNVMVDOTLTVGAENWRDKLDDP  
SSTSLVNDRLDISGSAADRSKNHSOISALYIEDNIEPVCTNII PGLRPDLYSD  
SGNFSPLNLSQBLGDFYKAGVARTFKAPNLYQSEGVLLYKNGNCPKDIITSGG  
CYLNGKDLDEPISVNEIGLETFWEDTHASVYFRNDYQNKIVAGDNIQOTASGAY  
ILKWQNGKALVDGIEASMSPLVKERLNMTNATMITSEQDKTGNPLSVIPKTYTN  
NSLNTWITQAFSAFNFTLYGKQPRTHAETRSDDTGLSGKELGAYSLVGTNPNYDI  
MKNLKLVGVSINLKNKIFRSSEGANVNEPRAYYAGVTASF"

BASE COUNT 686 a 557 c 632 g 620 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.44e-05 Length: 2495  
Score: 199.50 Matches: 84  
Percent Similarity: 41.94% Conservative: 59  
Best Local Similarity: 24.63% Mismatches: 135  
Query Match: 9.75% Indels: 63  
DB: 1 Gaps: 16

US-09-889-746-2 (1-400) x AFI35597 (1-2495)

Qy 9 TrpLeuSerLeuProLeuLeuSerValAlaValThrGlnGlnLeu-----Tyr 24  
Db 342 TGG---TCGCTAACTGTGCTCTAGTTGGTTGTAATAGCCAGGATCATGACCCAAATAC 398  
Qy 25 AlaGlnProAsnGluSerLeuProThrValGluLeuGluProValValIleThrIleAsp 44  
Db 399 TCCGACGATGATAATGAC-----GAGACTCTGTGTGGTGAA----- 434  
Qy 45 LysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrThrLysValIle 64  
Db 435 -----GCCACCGCTGAGCAGGATATTAAACACAGCAGCCGGCGTGTGCTGTTATT 482  
Qy 65 TyrGluGluGlnIleGlnGlnAlaThrGlySerArgGlnLeuAlaAspValMetAla 84  
Db 483 ACCACGAGGATATT---AAAAGACCCCTCCGGTAAACGACCTTTCAGATATTATTTCGT 539  
Qy 85 GlnLeuIleProSerLeuGlyVal-----SerSerGlyThrThrSerAsnPhe 100  
Db 540 AAAATG---CCTGGTGTAACTTACCGCAATAGCGCTCGGGCACACCGCGGTAAATAC 596  
Qy 101 GlyGlnThr-----MethIleGlyArgGlnValGlnPheLeuLeuAsnGlyVal 116  
Db 597 CGCCAGATCGATATTTCGTGGTATGGGGCGGAAACACCTTAATTTAATTGATGCTGTA 656  
Qy 117 ProLeuThr-----GlySerArgAspIleSerArg 126

Db 657 CCGGTGACGTACGTAACCTCGTGGTTATAGTGGCGTGGGAGCGTGATACCGCGGT 716  
Qy 127 GlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGly---Ala 145  
Db 717 GACACCAACTGGGTGCCACCGAACAGGTTGAGCGTATTGAAGTGATCCGGCGCCTGCG 776  
Qy 146 ThrSerIleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLysSerAsp 165  
Db 777 CGCGCGCGCTACGTTTCGGGGCGCGCGGGGGTGTGAACATCATTAACCAACAGTCCC 836  
Qy 166 LeuGluGluGlnPheGluThrArgIleGlyValHleGlySerLysValLeuSerSerGlu 185  
Db 837 ACCAACGACTGGCAGCGTTGCTGCTGTATACACCAACCGCGGAAAGTAGGCAAGAG 896  
Qy 186 GlyIleGlyTyrGlnValGlyGlnSerValAlaGlyValSerGluAsnGlyAsnValLeu 205  
Db 897 GGCCTACGCGTCCGCCCAATTTTCAGCTTAGTGGCTCTGGCTGGTGTGCTCTTACC 956  
Qy 206 AlaArgLeuAspValAspTyr---ArgThrThrGlyGlyAlaPheAspAlaAsnGlyLys 224  
Db 957 ACGGCTTTGTATGTTAACTGAATAAAACCGATGCTGACAGTTGGGATATTAAT- 1010  
Qy 225 ArgIleAlaProGluProAlaGlnThrAspLys-----GlnAspSerLysSer 240  
Db 1011 -----TCTCGGTGGTACGAAACCGACCGCGGCGATGAGGGGTACGTAAAC 1058  
Qy 241 LeuSerValAsnThrAsnValAspTrpGlnLeuAspLysGlnAsnIleAsnLeuAla 260  
Db 1059 AAGATATTAAACGGCTGTCTCGTGAAATTAATCCGACGACAGATTCCTCGATTTCGAA 1118  
Qy 261 LeuThrHisTyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeu 280  
Db 1119 GTCGGATAT-----AGCGCGCGGGAATATCTATGCGGCGCATACGCAGAACAGTTCT 1172  
Qy 281 AlaValLeuPheGlyGluLysProSerLeuAsnAlaIleLysGlyLeuSerLeuSerGlu 300  
Db 1173 TCCAGTGCAGTTACCGAA-----AGCTGGCAANA 1202  
Qy 301 GlnProLysThrThrLysSerThrPheAsnIleAsnTyr-----HisHisAspAspLeu 318  
Db 1203 TCCGCGAAAGACGACGACCGCTGTACCGACAGAAATTATGTCATTACGATAATGGTATC 1262  
Qy 319 Trp 319  
Db 1263 TGG 1265

RESULT 35  
SEU97227/c  
LOCUS  
DEFINITION  
Salmonella enterica ferric enterochelin esterase homolog (iroD),  
iroE (iroB) and TonB dependent outer membrane siderophore receptor  
protein (iroN) genes, complete cds.  
U97227  
U97227.1 GI:2738249  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Salmonella enterica.  
Salmonella enterica  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Salmonella.  
REFERENCE  
1 (bases 1 to 4837)  
Baumler, A.J., Norris, T.L., Lasco, T., Voight, W., Reisbrodt, R.,  
Rabsch, W. and Heffron, F.  
iron, a novel outer membrane siderophore receptor characteristic of  
Salmonella enterica  
J. Bacteriol. 180 (6), 1446-1453 (1998)  
98175667  
9515912  
REFERENCE  
2 (bases 1 to 4837)  
Baumler, A.J., Norris, T.L., Lasco, T., Reisbrodt, R., Rabsch, W. and  
Heffron, F.  
Direct Submission  
JOURNAL  
Submitted (12-APR-1997) Medical Microbiology & Immunology, Texas  
A&M University, 407 Reynolds Medical Building, College Station, TX

77843-1114, USA  
FEATURES  
Location/Qualifier  
source  
1..4837  
/organism="Salmonella enterica"  
/strain="A1870"  
/serotype="Typhi"  
/db\_xref="taxon:28901"  
36..1304  
/gene="iroD"  
36..1304  
/gene="iroD"  
/note="iroD"  
/codon\_start=1  
/transl\_table=11  
/product="ferric enterochelin esterase homolog"  
/protein\_id="AAC46181.1"  
/db\_xref="GI:2738250"  
/translation="MKLSPPVAPVAVIQTATRLRPLAAGSVDYDASHFRANSALPL  
VTAINGADIEREVTFLMRASPLRGVYVLAIVNTDKDVARDDATATDTHLTLRL  
PASVCGSYTMVEIPEKTPDETVALQLSRSLVGRADPLNSTPGINVRGNAESVIAL  
DLAPAEWESGVRLTSGSPONISGPGRGVRLVLPVPEQPLGLVLTDEGIMF  
HLGVSADIDAIIRSGRIAPVAVIAGVDINARERVALIGRRRLVLDIARLLPTLRA  
KYPROWADRTQTVLAGOSIGVTALMAARHAPESGLVLSHSPSWMTTPDRNRP  
PSAERSWVSEVLSNPSPAVTTHLCVSGLSGVTVQVAKLHEKLRAGVESHCCVYT  
GGHDYAMWVGALIDGLRLPRIKHLPRKI"  
1312..2229  
/gene="iroB"  
1312..2229  
/gene="iroB"  
/codon\_start=1  
/transl\_table=11  
/product="iroB"  
/protein\_id="AAC46182.1"  
/db\_xref="GI:2738251"  
/translation="MYGROYHNKCYRVAFVYFCSALSYAQPDMQPIGRPMADKGS  
GYHFRVNDPQSDGAKHRYVTAIPNKAPSPGIPVLTMLGNAMWDLPESTLQKL  
ADHSPPVIAIVGYQTLPPDLNGRAVDYPAQIDDDENNPRFRKKGPPAPARQL  
LERRHIAVQEQQITINPERGIIWHSYGLFVLDWSLSSPFIYISAPSLSDNFA  
LNRLLTAVKPSFPCHKLIIMEGSAENGSDRQWAMELLQKVQETVRLTLENGVNAAL  
GHYPGIHGPMNASFRSALLGISREPAS"  
2289..4472  
/gene="iron"  
complement(2289..4472)  
/function="receptor for 4-(3,4-dihydroxybenzyl)iden-amino  
benzoic acid"  
/note="iron"  
/codon\_start=1  
/transl\_table=11  
/product="TonB dependent outer membrane siderophore  
receptor protein"  
/protein\_id="AAC46183.1"  
/db\_xref="GI:2738252"  
/translation="MGMRVKKFTWLITVYSTGINSPLAABSTDNGETIVVESTAEQ  
VAKQPGVASIITRDDIQKNPVDNLADIIIRKMPVLTNSASGTGNRODIDIRWG  
PENTLVLDIGVPTSNVSVYRWGRDRDGTNWPVPMVRIEMIRGPAARVYSG  
AAGGVNIIITKRPNDMHGSLIYTYNPPSSKSGDTRRGNFSLGADDTSTMLY  
GNLRTDADSNWINSAGTKNAAGREGVAKDINSVRSKMTPOQILDEEAYSRQGN  
IYAGDTNNSNSAVYKSLAOSGERTRLRYNGYLTNKGSGRLSEFYFEKTDNT  
RMBGHSGGGERTINDQTFTNRLSYTSGVNVFVILFEQTLTVAEWNREDLN  
DPSSTLSLVKDSNIAGIPGSANRSSKNSEISALVEENIEMACTNIIPGLRPLY  
BSGGSFSPSLNSDELGEFVKVAGIAFAFAPNLYQSEGLVLSKNGCPKIDTS  
GGCYLVGNKLDPEISINKELIGLEFVDYDASVTFRYDQNKIYAGQOIGRSASG  
AVYLQWNGKALKLEIGIASMAVPLMDPDLNNNTNATWTSEOKDTGPNLSIPYT  
VNTPLMTITNALSNVNTLYGKKPRHNASRSETGLSGKLAGVSYGANVY  
DINKKRLAVGINSIFDKOITYSABGANTYNEPGRYAVGVNASF"  
misc\_binding  
complement(4558..4576)  
/note="Fur box"  
/bound\_molecule="Fur"  
BASE COUNT 1104 a 1330 c 1252 g 1151 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3..31e-05 Length: 4837

Score: 199.50 Matches: 89  
Percent Similarity: 40.47% Conservative: 66  
Best Local Similarity: 23.24% Mismatches: 135  
Query Match: 9.75% Indels: 94  
DB: 1 Gaps: 19  
US-09-889-746-2 (1-400) x SEU97227 (1-4837)  
QY 7 PheGlnTrpLeuSerLeuProLeuLeuSerValAlaValThrGlnGlnLeuTyraIle 25  
DB 4451 TTCATCTGGTTA---ATAACCGTGGTTTCTTACAGGAGTAAATGTCATTATACAGCA 4395  
QY 26 ---GlnProAsnGluSerLeuProThrValGluLeuGluProValIleThrIleAsp 44  
DB 4394 GAATCTACAGACGACAAACGGCGAAGCATAGTGGTTGA--- 4356  
QY 45 LysSerGlyMetAlaLeuAlaAsnAlaGlyIleThrGlnMetProHisThrThrLysValIle 64  
DB 4355 -----TCCACTCCGACGACAAAGTATTAAGCAACAGCCGGCGCTTCGATTATT 4308  
QY 65 TyrGluGluGlnIleGlnGluGlnAlaThrGlySerArgGlnLeuAlaAspValMetAla 84  
DB 4307 ACTCGGACGATTTACAAAGAACTCTCC---GTTAACGATCTTCCGATATTATTCGT 4251  
QY 85 GlnLeuIleProSerLeuGlyValSerSerGlyThrThrSerAsnPhelGlyInThrMet 104  
DB 4250 AAGATG---CTGGGGTTAATCTTACACCAATAGCGCTCG-----GGAGCGCGGA 4200  
QY 105 HisGlyArgGlnValGlnPhe-----LeuLeuAsn 114  
DB 4199 AATPACCGACCAATGATATTGCTGTGATAGGGCGGAAATACCTTAGCTGATTGAC 4140  
QY 115 GlyValProLeuThr-----GlySerArgAspIle 124  
DB 4139 GGGTCCCGGTGCGTCCGGAATTCGTTCCGTATAGCTGCGCGGAGAACCGCAC 4080  
QY 125 SerArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGly 144  
DB 4079 CGTGGGATACGAACCTGGGTGCGCGGAGAGTGTGAACGTATTAATGATCGGTGG 4020  
QY 145 ---AlaThrSerIleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLys 163  
DB 4019 CCGCGCGCGCGCGGTATGCTCCGGTGGCGGAGGGGTGTGAACATCAACACCA 3960  
QY 164 SerAspLeuGluGluGlnGlnPheGluThrArgIleGlyValHisGlySerLysLeuSer 183  
DB 3959 CGCCCACTAACGACTGGCATGGCTGCTGCTTATACCAACTACCCGGAACGAC 3900  
QY 184 SerGluGly-----IleGlyTyrGlnValGlyGlnSer 194  
DB 3899 AAGAAAGGTGACACCGCGGGGTAACTTAGCCTTAGCGGCCGTTGGCTGTGATACG 3840  
QY 195 ValAlaGlyValSerGluAsnGlyAsnValIleuAlaArgLeuAspValAspTyrArgThr 214  
DB 3839 TTATCCACATGCGCGCTGATCTAAT---CTCAACAGAACCATGCGCAC----- 3792  
QY 215 ThrGlyGlyAlaPheAspAlaAsnGly-----LysArgIleLeuLysProGluPro 230  
DB 3791 -----AGTGGGATTTAACTCTCCCGCGGTACCAAAATGCGCGGACGGAG 3741  
QY 231 AlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsnThrAsnValAspTyrGln 250  
DB 3740 GGGGTACCAATTA-----GATATTATAGCGCTTCTGATGGAAA 3699  
QY 251 LeuAspAspLysGlnAsnIleAsnLeuAlaLeuThrHisTyrAsnAspLysGlnAspThr 270  
DB 3698 ATGACCCCGGACCAATCTCTGATTTTGAAGCGGATAC-----AGTGGCAGCGGAAT 3645  
QY 271 AspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeuPheGlyLysLysProSerLeu 290  
DB 3644 ATCTATGCGCGGCGATACGCAAAAC-----AGTAATTCA 3612  
QY 291 AsnAlaIleLeuGlyLeuSerLeuSerGluGlnProLysThrThrLysSerThrPheAsn 310



```

/ds xref="GI:17429422"
/translation="MTNELQGLAKIARVFAHAALAAENRQIAEVEIRIQAETIR
TERDAMVADRSTVIAKTIERARQIRIQTSLDKPLTGSARQDLGLQPGASHPAPHPT
PHEQ"
/gene="RSC2402"
/feature="5032
4715..5032
/note="RSC2716"
4715..5032
/gene="RSC2402"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD16109.1"
/db_xref="GI:17429423"
/translation="MSSKQIRIVNIAIGAVKRVSPDNEALREAAAVDQMMIRRS
STTKIERNMAVAISISADLLAVORAAADALPIDAIRAKADIAARADEALRQVG
EAA"
/gene="6083
5577..6083
/gene="RSC2403"
/note="RSC2715"
5577..6083
/gene="RSC2403"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAD16110.1"
/db_xref="GI:17429424"
/translation="MLRTPTPTPAIMAAQVIMKSPDEASIDTLQTEVTLPMYGR
NYQANFMRRDRIRIDGVLPYHSSCPQGIAGLAEVCTSYPPDTQSDPSADPA
KOEIRFMVADRFVKCALIDLPTRAOEPLAEMTILQGNRSLITPPAPQRYITL
SLHLIHAD"
/gene="6081
6081..9793
/note="ACUR or RSO2714
predicted by Framed"
/evidence=not_experimental
6081..6846
/gene="RSC2404"
/note="RSC05747"
6181..6846
/gene="RSC2404"
/function="miscellaneous; hypothetical/partial homology"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon usage
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL SIGNAL PEPTIDE PROTEIN"
/protein_id="CAD16111.1"
/db_xref="GI:17429425"
/translation="MKPAALVLTGATVAVATSAALQTPSAAGMTPEPSGLSIDAQA
VAEVPDLYTLTGAEQADGASISLSRAEDVIAQAKRTAGVQAEYTGFPVYEN
TRNGKINVMRRAEVRITSKDPAASLAGIAGQVQVNVFLISEATAETEL
VLTPCROGQGAASSKAQKFAFERQEPPLIDTLRLIQVGNRRCQAAPRNCPE
TTR"
complement(6668..7901)
/note="predicted by Homology"
repeat_region

```

Pred. No.:	0.00474	Length:	20450
Score:	198.00	Matches:	105
Percent Similarity:	39.77%	Conservative:	66
Best Local Similarity:	24.42%	Mismatches:	163
Query Match:	9.67%	Indels:	99
DB:	1	Gaps:	18

  

US-09-889-746-2 (1-400) x AL646070 (1-204050)	
QY 31 LeuProThryValGluLeuGluProValValIleThrIleAspLysSerGlyMetAlaLeu 50	DB 3780 ATGCTGTCGGCGGAATCAACCGGACCGGATGGAGCGGCTGCGGAGGAGCAAGCTT 3721
QY 51 AlaAsnArgIleThrGlnMetProHISThrIlyValIleTyrgluGluGlnIleGln 70	DB 3720 GCCCAT-----GGCGTCGGCAATACCAGCGATCATCTCGCGCGCGACATCGAGGG 3670
QY 71 GluGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeu---IleProSer 89	DB 3669 TCGGAGGCAACCGAGTGGGATGGCTGCTGCGCGCGGAGGCGGACATGAGATGCCAG 3610
QY 90 LeuGlyValSerSerGlyThrThrSerAsnPhelGlyGlnThrMetHisGlyArgGlnVal 109	DB 3609 GCGGGCGGCGCGGCGCGCGCGGACCTTCTATCGCGGCGCGCGGATCGACACAG 3550
QY 110 GlnPheLeuLeuGlnGlyValIProLeuThrIlySerArgAspIleSerArgGlnLeuAsn 129	DB 3549 CTCATTTCTGATCGACGCGCGCGCGGAGATTCGCGACACACGAGACAGATCGAG 3490
QY 130 SerIleAsnProAsnGlnValAlaArgIleGlyValLeuSerGly---AlaThrSerIle 148	DB 3489 CAGCTGATGGCCCATCAAGATCAACCATATCGAGATGTGCGCGCAATGTCTCGCGCTG 3430
QY 149 TyrglySerGlyAlaThrGlyIleuIleAsnIleValThrLysSerAspLeuGlnIu 168	DB 3429 TACGCGTCGATGGATCGCGCGCGGATCGAGATTTTACACCGCAACGCGCGGAC 3370
QY 169 GluGlnPhe-----GluThrArgIleGlyValHisGlySerLysLeuSerSerIu 185	DB 3369 GCGCGCTGGCCAAACGCGGAATCGAGTACGCGCGCGGACATACAGACGCGGAGGC 3310
QY 186 GlyIleIleGlyTyrglnValGlyGlnSerValAlaGlyValSerGluAsnGlyAsnValIleu 205	DB 3309 GGCATC-----AGCGGACGACTGGAC 3289
QY 206 AlaArgLeuAspValAspTyrr-----ArgThrThrGlyGlyAla 218	DB 3288 GCGCGCGGCGCAACGCTGTTGCGCGCTGCGGATTCGCGCTTAAGACGACGGGT----- 3239
QY 219 PheAspAlaAsnGlyValAspGlnIleAlaPro-----GluProAlaGlnThr 233	DB 3234 TTCCTGTCGATCAATCCGAAGACGGCGCGGACGCGGACCCGACGACGACGACGCTGAGCGG 3175
QY 234 AspLysGlnAspSerLysSerLeuSerValAsnThrAsnValAspTrpGln----- 250	DB 3174 AACAAAGCGGTGCGCGGACGCTGACGATCGCTTCTGTCGACGCTGAGCGCGGCTG 3115
QY 251-----LeuAsp 252	DB 3114 ACGGTGTTCCAGACCTGAGACACCGTAGACTACGACAAAGCGCTTGCGCGCGCCACCGAC 3055
QY 253 AspLysGlnAsnIleAsn-----LeuAlaLeuThrHisTyrglnAspLysGlnAspArg 270	DB 3054 GACCAACGATGCGACACGAGCGGTGCTGATCGATCGGCTGAGTGAAGGACGACGCTGACG 2995
QY 271 AspTyrglnAspTyrglyAsnArgLeuAlaValLeuPheGlyGlyLysAspProSerLeu 290	DB 2994-----CCGACTGAGAGACGCGCTGACGCTGTCGACGAGGCGACGACAAAGACTGG 2944
QY 291 AsnAlaIleLysGlyLeuSerLeuSerGluGlnProLysThrThrLysSer----- 307	DB 2943 AATTTCACCAACGCGCGATGTCAGAGACCGGCGCGCTTCACACGCGGACCGGACCGG 2884



```

/codon_start=1
/transl_table=11
/product="Y1094 protein"
/protein_id="CAC43411.1"
/db_xref="GI:14594863"
/db_xref="SPTREMBL:P74994"
/translacion="MELQHORLMLALQQLQESLISAAPALSQQAVDQESYMDLE
HLHEBKARHQKQAMYTMAAFAVITFEEDPTFATGAQOKQLSLSTERN
ENIVLLGSVGTALAMGYEAVRAGIYVFTADLLQLSTAQKQKXTLOR
GVMAFRLIIDEIGLIPSOEAKLFFQIARVYKSMILUSNLPFGQMDTFAGDA
ALTSAMDRLIHLHSVVOIKGSYRLRQKRAVIAENPE"
/complement(4062..5813)
/feature="ORF1"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC43412.1"
/db_xref="GI:14594864"
/db_xref="SPTREMBL:Q93K86"
/translacion="MYPLIHGIIYCDTLISHPOKHLPRKLTAGOSFAGIAMQYH
VTDIQRHLHLRFPVNNQPHADNLOGHPDAICSADEDOYGFIIATQIAVAKGI
GREGAVIGIEGPGSGKSLINLRMLVBOIECTPLVITISPLDGSNTSLVSLI
LPVANIITAEERQRLAPERRALRRKSLTRAKTMDITPRATARNLATVAAAVIP
GVPDGALKALSETRMLEKKTIAEMKTEIAKKIDELDSFVLLDLDLBEAQA
VEVIRLVSVDGPFPRRYLLCYDKAILSQALSGVPGNLYLQKIVQISFLRPR
TFVLRKRPRDAALYRKVNGHAPRIYEQULQVADYGALKTPREVQVYMLNLTPE
LYPGMDVYVPDLCELOLRTTNAGLYDMVEYSEBAVVAAGGVSEBQAMAS
SLKSHLARYPAFAHVAHVAHVAHVAHVAHVAHVAHVAHVAHVAHVAHVAHVAH
QAVNYLYPAFAHVAHVAHVAHVAHVAHVAHVAHVAHVAHVAHVAHVAHVAH
HILQVLYQVILTFVNDG"
/feature="ORF2"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC43413.1"
/db_xref="GI:14594865"
/db_xref="SPTREMBL:Q93K85"
/translacion="MOSVSKSIYRGKLSLPAVGEIHKEIVASEETVWISLOREIT
ANILGKDPRLIVIGPSIHVOAAYEAKLSVLONKYSOMYIWRVTEKPRTR
KMKGIMHBPDLNGSYNEKGRYARQCSITTRVATFELPILTPYADICW
GAVGARITBSQTHROLASGLCPVGRKNSIDNIALDAIILAAEQHVAWTSLTKC
ISTLLTDNPHGLILRGREPNYGLSDITKAVKLMHDEGINHRLIIDCSHNSGKVA
ERQISVAREVINDRKMFGYVAGIMLESFLQGGKQSDSLPREYGSVTDECLSMQTE
QULSTLAQL"
/insertion_beg="10458"
/insertion_end="10458"
/complement(9245..10099)
/feature="trp"
/complement(9245..10099)
/feature="trp"
/function="transposase-encoding"
/codon_start=1
/transl_table=11
/product="Tnp protein"
/protein_id="CAC43414.1"
/db_xref="GI:14594866"
/db_xref="SPTREMBL:Q93K84"
/translacion="MKYVFLENHRAFSIKAMCRVLVARSQVYVWLRHHQMSLRQO
FRULTDAVHKAFKAKORYGAARLADENPEFNITIASLRQGLRAKASRSPS
YRAHGLPVLENLEODFSASGPNKAGIITLRTDEGLYLAVIDLMSRAVISM
SPRMTAOLACDALQALWRRRRPESVITHTDGGYCSGDYQALLKRHLRGSMAKG
YCYDNACVESPFHSLKVCIEHGRFSRIMATYFNTIECDYNNRRHRSACGSLSP
OPENHMLA"
/insertion_beg="10491"
/insertion_end="10491"
/complement(11546..10491)
/feature="cvab"
/feature="cvab"
/function="ATP-binding protein"
/codon_start=1
/transl_table=11
/product="cvab protein"
/protein_id="CAC43415.1"

```

```

/db_xref="GI:14594867"
/translacion="MELQHORLMLALQQLQESLISAAPALSQQAVDQESYMDLE
HLHEBKARHQKQAMYTMAAFAVITFEEDPTFATGAQOKQLSLSTERN
ENIVLLGSVGTALAMGYEAVRAGIYVFTADLLQLSTAQKQKXTLOR
GVMAFRLIIDEIGLIPSOEAKLFFQIARVYKSMILUSNLPFGQMDTFAGDA
ALTSAMDRLIHLHSVVOIKGSYRLRQKRAVIAENPE"
/complement(4062..5813)
/feature="ORF1"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC43412.1"
/db_xref="GI:14594864"
/db_xref="SPTREMBL:Q93K86"
/translacion="MYPLIHGIIYCDTLISHPOKHLPRKLTAGOSFAGIAMQYH
VTDIQRHLHLRFPVNNQPHADNLOGHPDAICSADEDOYGFIIATQIAVAKGI
GREGAVIGIEGPGSGKSLINLRMLVBOIECTPLVITISPLDGSNTSLVSLI
LPVANIITAEERQRLAPERRALRRKSLTRAKTMDITPRATARNLATVAAAVIP
GVPDGALKALSETRMLEKKTIAEMKTEIAKKIDELDSFVLLDLDLBEAQA
VEVIRLVSVDGPFPRRYLLCYDKAILSQALSGVPGNLYLQKIVQISFLRPR
TFVLRKRPRDAALYRKVNGHAPRIYEQULQVADYGALKTPREVQVYMLNLTPE
LYPGMDVYVPDLCELOLRTTNAGLYDMVEYSEBAVVAAGGVSEBQAMAS
SLKSHLARYPAFAHVAHVAHVAHVAHVAHVAHVAHVAHVAHVAHVAHVAHVAH
QAVNYLYPAFAHVAHVAHVAHVAHVAHVAHVAHVAHVAHVAHVAHVAHVAH
HILQVLYQVILTFVNDG"
/feature="ORF2"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAC43413.1"
/db_xref="GI:14594865"
/db_xref="SPTREMBL:Q93K85"
/translacion="MOSVSKSIYRGKLSLPAVGEIHKEIVASEETVWISLOREIT
ANILGKDPRLIVIGPSIHVOAAYEAKLSVLONKYSOMYIWRVTEKPRTR
KMKGIMHBPDLNGSYNEKGRYARQCSITTRVATFELPILTPYADICW
GAVGARITBSQTHROLASGLCPVGRKNSIDNIALDAIILAAEQHVAWTSLTKC
ISTLLTDNPHGLILRGREPNYGLSDITKAVKLMHDEGINHRLIIDCSHNSGKVA
ERQISVAREVINDRKMFGYVAGIMLESFLQGGKQSDSLPREYGSVTDECLSMQTE
QULSTLAQL"
/insertion_beg="10458"
/insertion_end="10458"
/complement(9245..10099)
/feature="trp"
/complement(9245..10099)
/feature="trp"
/function="transposase-encoding"
/codon_start=1
/transl_table=11
/product="Tnp protein"
/protein_id="CAC43414.1"
/db_xref="GI:14594866"
/db_xref="SPTREMBL:Q93K84"
/translacion="MKYVFLENHRAFSIKAMCRVLVARSQVYVWLRHHQMSLRQO
FRULTDAVHKAFKAKORYGAARLADENPEFNITIASLRQGLRAKASRSPS
YRAHGLPVLENLEODFSASGPNKAGIITLRTDEGLYLAVIDLMSRAVISM
SPRMTAOLACDALQALWRRRRPESVITHTDGGYCSGDYQALLKRHLRGSMAKG
YCYDNACVESPFHSLKVCIEHGRFSRIMATYFNTIECDYNNRRHRSACGSLSP
OPENHMLA"
/insertion_beg="10491"
/insertion_end="10491"
/complement(11546..10491)
/feature="cvab"
/feature="cvab"
/function="ATP-binding protein"
/codon_start=1
/transl_table=11
/product="cvab protein"
/protein_id="CAC43415.1"

```







regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

# FEATURES

```

source
    Location/Qualifiers
        1..20513
            /organism="Salmonella typhimurium LT2"
            /strain="LT2; SGSC 1412; ATCC 700720"
            /db_xref="ATCC:700720"
            /db_xref="taxon:99287"
            /note="LT2"
            complement(221..688)
            /gene="STM2764"
            complement(221..688)
            /gene="STM2764"
            /codon_start=1
            /transl_table=11
            /product="putative integrase core domain protein"
            /protein_id="AAL21650.1"
            /db_xref="GI:16421312"
            /translation="MQAAERYPRYGFKLQVLRROCYMMNHKRIHRIYCLLKLDLFR
            RKGORLPVRNPSPLATPEALNQSWSDPMHDALVCGRRPFTFNVDNREALSIEI
            DLNLPAIRVVLVRIANRCYPVMLRMDNGPEISLALAEWAQHAVKLEFI"
            complement(835..1108)
            /gene="STM2765"
            complement(835..1101)
            /gene="STM2765"
            /codon_start=1
            /transl_table=11
            /product="putative transposase"
            /protein_id="AAL21651.1"
            /db_xref="GI:16421313"
            /translation="MRKARFTEHQIIVIKSVKAGRTVKDCAKEAGISEATYYNWKSR
            YGMEPSDIKKIKOLEDNRLKQMFADLSLENALKDVIKKL"
            complement(1103..1108)
            /gene="STM2765"
            /note="putative RBS for STM2765; RegulonDB:STMS1H002840"
            complement(1554..2199)
            /gene="STM2766"
            complement(1554..2192)
            /gene="STM2766"
            /codon_start=1
            /transl_table=11
            /product="putative cytoplasmic protein"
            /protein_id="AAL21652.1"
            /db_xref="GI:16421314"
            /translation="WSLYFPNNVPAEYLDKFCVARDAPSNLENLLVTAEIILNTCHESW
            NKETKDFLLSTGTHKRILLYKKTGDFTNLPRQVIEYESNICFNVDAYALPVNAEP
            ISRCNRVIATCGNSFSYEALAVELCDFRDIQAINICDAISLLLDVHGYPFRDD
            DLKARGKVPYHFDPCNNSTNVKGSNIRIGDTPLDLFDVSKORPYLT"
            complement(2189..4178)
            /gene="STM2767"
            complement(2189..4171)
            /gene="STM2767"
            /codon_start=1
            /transl_table=11
            /product="putative superfamily I DNA and RNA helicase"
            /protein_id="AAL21653.1"
            /db_xref="GI:16421315"
            /translation="MSNYCFISQDALALQAAGVDVINSYAEQHKQTYILCRPLSN
            EDVYDRAIAVFSFGIKPFPIFDGDDDLFEYQEDFLEDVSYLAEKFKYRKIGR
            KSKWQILPESLRNDIDFKLEVTKEKSRVIDLIISLIVGINDTSINLEANNLLDT
            IKSIIILPDTQTKPFDQSGFKGKSVIOGLAGSGKTELLHLKLEIVSKNPDNRIAPT
            CFNKILASTMTIRIPERDFMRVEKQIEWGTLCFCNSWGUTKBPFGSMGYVICHYE
            IPFGGNDGFDALCKKALADINNSGRADKALDYVFDDESQDFQSIDLCCEWYTSK
            KLYAGDVQFQIFPISNVNRADIVLKCYRTDPKMLFHALCMGLYEFPVLRWLK
            EPWSDSCYKTKYKGDVRVHLSRDPURREFIPKNHKSTAVHLLLEGTONGPKIVDIII
            DIKERNPSELEGDTAVIFLDAGGYIYEIHSKXKVKQQLGWDNSHETKSKQDGKL
            FISINNAKLEFPFVFCFAMKLVKRAFNRLYTMARSPFLESHLVNLNDNENPAIP
            TILGFLNPLNNYMDVRLPSPDEIQTOSQKDFIVLDESVSISQMVKSYCADKSKTPRLI
            AKITDVERIITAEODDADGEVILKGLIEIYEYENKKL"
            complement(2194..2199)

```

```

RBS
    /gene="STM2766"
    /note="putative RBS for STM2766; RegulonDB:STMS1H002841"
    complement(4173..4178)
    /gene="STM2767"
    /note="putative RBS for STM2767; RegulonDB:STMS1H002842"
    4438..4749
    /gene="STM2768"
    4438..4443
    /gene="STM2768"
    /note="putative RBS for STM2768; RegulonDB:STMS1H002843"
    4450..4749
    /gene="STM2768"
    /note="similar to E. coli orf, hypothetical protein
    (AAC75149.1); Blastp hit to AAC75149.1 (102 aa), 85%
    identity in aa 4 - 102"
    /codon_start=1
    /transl_table=11
    /product="putative transposase"
    /protein_id="AAL21654.1"
    /db_xref="GI:16421316"
    /translation="MTKPALTTKPKRQHTPEFRQEAALKAKRIGVAAAARELSLYKS
    QLNWRSKQONQLSSSREGEQSAETARLKRQLAERDELAIIQKAATYPAKRLK"
    4732..5612
    /gene="STM2769"
    4732..4737
    /gene="STM2769"
    /note="putative RBS for STM2769; RegulonDB:STMS1H002844"
    4746..5612
    /gene="STM2769"
    /note="similar to E. coli IS3 putative transposase
    (AAC75150.1); Blastp hit to AAC75150.1 (288 aa), 84%
    identity in aa 1 - 288"
    /codon_start=1
    /transl_table=11
    /product="putative transposase"
    /protein_id="AAL21655.1"
    /db_xref="GI:16421317"
    /translation="MKYVFIEHQHAEFNKAMCRVQFVARSGWYVHQRHQRINROR
    FRVCDNVVREAFSDAKQRYGAPRLTDELQAGYQFNVTAAASLRQGLRAKASRRP
    RPVSYRKHGLPVSENLKQDFYASGPNQKWGDIITYRTGEGMLYLAVYLDLWPRSVI
    GWSMSRMTAQLACNALQMALQWKCPENIVHTDRGOYCSYDQYSLKRNLRGSM
    SARGCCYDNACAESFHTLVECIHGDFVSRIMRTAVFNYSCECINRWRHRSACGG
    LSPQEFENQLA"
    complement(6392..6931)
    /gene="fljA"
    /note="STM2770"
    complement(6392..6931)
    /gene="fljA"
    /note="flagellar synthesis; repressor of phase-1 flagellin
    gene. (SW:FLJ_A_SALTY)"
    /codon_start=1
    /transl_table=11
    /product="repressor of fliC"
    /protein_id="AAL21656.1"
    /db_xref="GI:16421318"
    /translation="MECMAYNDISYGREAEIWPDRYSMLARRVQFLRFNDIPVRLVSN
    NARIITGYIAKPNPKENILASDPKGNKRIEVLKSLAILELGSNDANFLSLVPAD
    HMVQYPTPSRRDYFSICNKYQKGVGKIYMKYQVLTGKTGTGNACQGVRTSNGN
    HMVQYPTPSRRDYFSICNKYQKGVGKIYMKYQVLTGKTGTGNACQGVRTSNGN"
    complement(6999..8526)
    /gene="fljB"
    /note="STM2771"
    complement(6999..8519)
    /gene="fljB"
    /note="phase 2 flagellin; flagellar synthesis; phase-2
    flagellin. (SW:FLJ_B_SALTY)"
    /codon_start=1
    /transl_table=11
    /product="filament structural protein"
    /protein_id="AAL21657.1"
    /db_xref="GI:16421319"
    /translation="MAQVINTNSLSLLTQNNLNKSQSALGTAIERLSSGLRINSKOD
    AAGQAIAIRRTANIKGTQASRRANDGISIAQTTEGALNINNLRVRELAVOSANS"

```

TNSGSDLSIQAEITQRLEIDRVSGQTFQNGVKVLAQDNTLTIOVANGDETIDIL  
 KQINSQTLGLDLSIQAEITQRLEIDRVSGQTFQNGVKVLAQDNTLTIOVANGDETIDIL  
 ASVTGCAVADADNNKYEVLTIGFTGADAKKGDYEVNATGTGTLAAGATKTPA  
 GATKTEVOELKDTPAVVSADAKNLIAGVDATDANGNELVMASITDNGKTIISGKY  
 AKADGKYAADYDEATGAIRAKTISYTPADTGTAAQNGDNGKREVTITDKGY  
 NASKAAGHDFKQPELEAKAATTEPLOTIDALAAQVDALSDJCAVONRNSAITN  
 LGNTVNNISEASRIEDSDYATSEVSNMSRAQLIQAGTSVLAAQNVQNVLSILR"  
 complement (8521.. 8526)  
 /gene="fljB"

## Alignment Scores:

Pred. No.: 0.000519 Length: 20513  
 Score: 194.00 Matches: 87  
 Percent Similarity: 40.53% Conservative: 65  
 Best Local Similarity: 23.20% Mismatches: 145  
 Query Match: 9.48% Indels: 79  
 DB: 1 Gaps: 17

US-09-889-746-2 (1-400) x AE008826 (1-20513)

Qy 7 PheGlnTrpLeuSerLeuProLeuLeuSerValAlaValThrGlnGlnLeuTyrAla--- 25  
 Db 19517 TTGATCTGGTTA---ATACCGGTGTTTCTACAGGGGTTAATGTCATTATCAGCAGCA 19461  
 Qy 26 ---GlnProAsnGlnSerLeuProThrValGlnLeuGlnProValAlaIleThrIleAsp 44  
 Db 19460 GAATCTACTGACAGCAACCGCGAAGCATGTGTGTTGAA----- 19422  
 Qy 45 LysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHISThrIlyValIle 64  
 Db 19421 -----TCCACCGCGGAGCAGATTTAAAGCAACAGCGCGGCGTTTCGATTATT 19374  
 Qy 65 TyrGlnGlnGlnIleGlnGlnGlnAlaThrGlySerArgGlnLeuAlaAspValMetAla 84  
 Db 19373 ACTCGTACGACATTTCAAAAGATCCTCC---GTTACACATCTTCCGCAATTATTTCGT 19317  
 Qy 85 GlnLeuIleProSerLeuGlnValSerSerGlyThrThrSerAsnPhenylGlnThrMet 104  
 Db 19316 AAGATG---CCTGGGGTTAATCTTACCGGCAATAGCCCTCG-----GGACGCGCGCA 19266  
 Qy 105 HisGlyArgGlnValGlnPhe-----LeuLeuAsn 114  
 Db 19265 AATTAACCGAATAATTGATTCTGTGTATGGGCGGAAATACTTATGTCGATTGAC 19206  
 Qy 115 GlnValProLeuThr-----GlySerArgAspIle 124  
 Db 19205 GGGGTGCGGTGACGTCGCCAAATTCGCTTCCTATATGCTGCGCGGAGAAACCGACACC 19146  
 Qy 125 SerArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGlnValLeuSerGly 144  
 Db 19145 CGTGGGATACGAACTGCGGATACCGCGAGATGTACAGTATGTAAGTATCCGTGG 19086  
 Qy 145 ---AlaThrSerIleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLys 163  
 Db 19085 CCGCGCGTCCGCGCTATGCTCCGCGCGCTGCGCGGTGTGATATCATCACCAGAA 19026  
 Qy 164 SerAspLeuGlnGlnGlnGlnPheGlnThrArgIleGlyValHisGlySerLysLeuSer 183  
 Db 19025 CGTCCGACTACGACTGCGACGCGCTCGCTGCTGTATATCCAAACAGCCGAGAAAGCAGC 18966  
 Qy 184 SerGlnGlyIleGlyTyrGlnValGlnSerValAlaGlyValSerGluAsnGlyAsn 203  
 Db 18965 AAGAAAGGTACACCGCGGGGGGTAATTACTTACCGGTGCGGTGGTGATACG 18906  
 Qy 204 ValLeuAlaArgLeuAspValAspTyr---ArgThrThrGlyGlyAlaPheAspAlaAsn 222  
 Db 18905 TTACAGCATGCGCTGTACCGTATCTCAACAGAACCCATCCGATCGGTGATATTAAAC 18846  
 Qy 223 Gly-----LysArgIleAlaProGlnProAlaGlnThrAspLysGlnAspSer 238  
 Db 18845 TCCCTCCGCGCGTACCAAAATGCCCGCGAGCGGAGGGGGGTACCAATTA--- 18795  
 Qy 239 LysSerLeuSerValAsnThrAsnValAspTyrPheLeuAspAspLysGlnAsnIleAsn 258

Db 18794 -----GATATTATATGCGCTTCTCATGGAATAATGACCCCGAGCAAAATCCTTGAT 18744  
 Qy 259 LeuAlaLeuThrHisTyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsn 278  
 Db 18743 TTGGAACCGGATAT-----ACTCGGAGGGGAAATATTATTAAGTCCGATACCGCAAAAC 18690  
 Qy 279 ArgLeuAlaValLeuPheGlyGlnLysProSerLeuAsnAlaIleLysGlyLeuSerLeu 298  
 Db 18689 AGTACTTCA-----AACGCGGTAC---AAAGCTCTG 18660  
 Qy 299 SerGlnGlnProLysThrThrLysSerThrPheAsnIleAsnTyr-----HisIleAsp 316  
 Db 18659 GCGCAGTCCGCGGCGGAGACCAATCCGCTGTACCGGAGAAATTACCGTCACTCAATAC 18600  
 Qy 317 AspLeuTyr-----GlyAsnThr 332  
 Db 18599 GGCATCTGGACTGGGGGCAACCGCTGGGGTCTTACTATGAAAAAACCGATTAACACC 18540  
 Qy 323 IleAsnThrAsnAlaTyrTyrArgArgGlnLysGlyArgPheTyr 337  
 Db 18539 CGCATGAATGAGGGGCT-CTCCGCGCGCGGCGGAGGGCGGTATTAC 18496  
 RESULT 40  
 AE012418 10118 bp DNA linear BCT 23-MAY-2002  
 LOCUS Xanthomonas campestris pv. campestris str. ATCC 33913, section 326  
 DEFINITION of 460 of the complete genome.  
 ACCESSION AE012418 AE008922  
 VERSION AE012418.1 GI:21114258  
 KEYWORDS  
 SOURCE Xanthomonas campestris pv. campestris str. ATCC 33913.  
 ORGANISM Xanthomonas campestris pv. campestris str. ATCC 33913  
 Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 Xanthomonas.  
 REFERENCE  
 1 (bases 1 to 10118)  
 da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,  
 Queglio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida  
 Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bercolini,M.C.,  
 Camargo,L.B.A., Camarotte,G., Cannavan,F., Cardozo,J.,  
 Chamego,F., Ciapina,L.P., Clocatelli,R.M.B., Coutinho,L.L.,  
 Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,  
 Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,  
 Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite  
 Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,B.C., Machado,M.A.,  
 Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,  
 Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinoia,L.A.F.,  
 Takita,M.A., Tamura,R.B., Teixeira,B.C., Tezara,R.I.D., Trindade dos  
 Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and  
 Kitajima,J.P.  
 Comparison of the genomes of two Xanthomonas pathogens with  
 differing host specificities  
 Nature 417 (6887), 459-463 (2002)  
 JOURNAL MEDLINE 22022145  
 PUBMED 12024217  
 REFERENCE  
 2 (bases 1 to 10118)  
 da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R.,  
 Queglio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida  
 Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bercolini,M.C.,  
 Camargo,L.B.A., Camarotte,G., Cannavan,F., Cardozo,J.,  
 Chamego,F., Ciapina,L.P., Clocatelli,R.M.B., Coutinho,L.L.,  
 Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,  
 Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,  
 Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite  
 Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,B.C., Machado,M.A.,  
 Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A.,  
 Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spinoia,L.A.F.,  
 Takita,M.A., Tamura,R.B., Teixeira,B.C., Tezara,R.I.D., Trindade dos  
 Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and  
 Kitajima,J.P.

Kitajima,J.P.  
Direct Submission  
Submitted (28-NOV-2001) Departamento de Bioquimica, Universidade de  
Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,  
Brazil

FEATURES  
source  
1..10118  
/organism="Xanthomonas campestris pv. campestris str. ATCC  
33913"  
/strain="ATCC 33913"  
/db\_xref="ATCC:33913"  
/db\_xref="taxon:190485"  
/note="pathovar: campestris"  
139..534  
/gene="XCC3044"  
139..534  
/gene="XCC3044"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAM42315.1"  
/db\_xref="GI:21114259"  
/translation="MIDHLHLVQDLKASRRFYCAVLDVLCVPLGGSGEGFFWADLP  
VSDRLVAQGLTGRHLAFQAHVDMVHFAAALAHAGGRDNGAPVRYPHPSYIAA  
FVLPDGNNEIAVPHGPAQSAEAVQITY"  
777..3197  
/gene="bfea"  
/note="XCC3045"  
777..3197  
/gene="bfea"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=11  
/product="ferric enterobactin receptor"  
/protein\_id="AAM42316.1"  
/db\_xref="GI:21114260"  
/translation="MHFPPTSRCHRMNCKTSLAVAVVLAFLSPASTAVAQSQAPTK  
SDGDAVPAQRLGLSDPQVLAIVNGKHYHSTALINDTOGRSSPADLNTIPAAV  
ERIEVRDCASQAGSDAIAGINIVILKSGEGSGVNGRYKYSAGDGEQVQLSGDAG  
PFACTGVHIAAAGQSDQNRALPFGQVEIKGDIPEIDQAVSNGEYSPTDILT  
PYSQVMSRVELSNGYRFAGDNRNRPPIPDGLPQIYNVSKDVSWGGLKSTEG  
GLNTDLSYNGQNNLTFOVRNSLNSGLASPTDFHAGTLEVTQNLVNAADPTKLDWG  
LAYVALAFAEAGKEGKFNQSPDAASANGGIPSAAGLIPGAQVPGFKLSAGVY  
NRNHSVAVDLEADLTEKLSAGLAKRYEKYSDFDGTATGKLSLYALTDKVALRATAS  
TGRAPSLQQQNFQSIATQPLNVAQPNQGVSAIPFEIGTFRTDNPALALGABEPKAB  
ESKNYGLVLOPVENLYITVDAYRIDIDDRIVLSENLTSAARNYLOANPPGPIGG  
RYFTNAVDTKTQGVGVGYRNMDSGVELTGTGYNNKTEVEQIAANPPALEAIDPG  
AVRIGRAELGRITGTGPRDFLAGTSPGNSFTGTATRWGEFTTGTNAGSDQTYA  
AKWTLDLAASKLGSWNFTVCGDNVLENYPDQEAAGLGRTRYLYPYSSASPFGFNGALV  
YANNYK"  
3675..5927  
/gene="bfea"  
/note="XCC3046"  
3675..5927  
/gene="bfea"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=11  
/product="ferric enterobactin receptor"  
/protein\_id="AAM42317.1"  
/db\_xref="GI:21114261"  
/translation="MAESESPIIDITEQSLQATGATATATGKLLPSLNPFRPAISD  
GNDARPAATRLGSDAVLVLDGKRYHTSALINPNYVGRGSPADNLSLNPMSAVAR  
IEVLDRGSAQYGDADAGVNNIVLKHGAKTGSDNSVSVYGMFMDQGEQNGIDSGV  
MGFGCTGGDTTPGWRLSNWNTQMTNRGENTRATVPPTPNPGVPYVYRHGDPKS  
TFYQGLAAGFYEFSPLELYGHFSASRREVTNGYRAADNPARNVPSIYPNGPLPQI  
YNPNVDRSAVLIGKSTSESNWDLASVYGNKDMSENLINSINTNLRYTTGASPTNEN  
SGFGETQGTQNLNIDISRAFDNGLAYPVNVAFGAEHRQDRYRITAGSPDSYFFDNTIN

PADGAPYBGCAOVFSGLEPSVACKFGRHSNAVYANLEANLTELKSGIAGRYENYSDA  
GATRSKLSRYAFTDTFSLRGITISNGFRAPSLAQNYASVVTILNGELVQVGTYRT  
SDVVALAGARPLSPKSSNYGVWQPTANFTSSLDVYQIRIMQILYSDQLAQ  
PIQQAQVFNVAQSRTRGVDMVNSYIADLGELQTLGASANYNKTITLLESNPN  
FGASQGLTDATPRTKYNASADWAIGGFTLGNATRYGSIKRISDPVDSQDQTYDA  
RWTLNLSASQTNNAFTFTVGADNLTNQYPTKQLTAYDDRAGGLQYSSLSPPGFNGR  
YWYGRVYRP"  
complement (6030..7223)  
/gene="bioi"  
/note="XCC3047"  
complement (6030..7223)  
/gene="bioi"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=11  
/product="cytochrome P450 hydroxylase"  
/protein\_id="AAM42318.1"  
/db\_xref="GI:21114262"  
/translation="MQLSDPATPAFRQDPYPMYARLARAGPLVOISNGWVSGHYTVV  
DALLSDRRVGRNYLDSIRVRYGANAEMFLQGMGRMFLNLPVHTQORALMTKAFG  
ARQLEALREVAVDTADALDQHRSDCLLNDPAMPMTISLIRMLGLADTVDAALG  
QASSALAKVPDPLMRPDMQAATAVYTLQYFRAIVLQRRDQEDDLIARLIAAEQH  
GQMPVDDIVSNVIMLFTAGHETTANNICNALIHRHPEQLQLLRDPTPTLMPNAVLE  
CMRYDSQVQVAMRSVLOPQVEGTTLPVGAILYMLGSAHNDRAEQFTAPQQLDLRRQ  
GRALSQGGVHHGCLGNKLALIELETALERLULQAPALRULPELONLSWNERANURGIQA  
LHATW"  
8295..9992  
/gene="ybaI"  
/note="XCC3048"  
8295..9992  
/gene="ybaI"  
/note="identified by sequence similarity; putative; ORF  
located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=11  
/product="cation:proton antiporter"  
/protein\_id="AAM42319.1"  
/db\_xref="GI:21114263"  
/translation="MHDTSLDIIIVAGLAVFVLGTQAKVKLSPLVGLVLAGVCVG  
VPTPGFVADQTMANQSELQVGLMLFGVHLFSLDDEMEVKIATVPALCAQIVVATLL  
GWLANSWGPLMQGLVFGALGVASTVLLRLEERLLETORGRITAVGMLIVEDILV  
MYATVLLPALANVLGSSAGAAHAGESTSLAALCHWTLKMYAFVAVMLVGRVYIQ  
WSKEVAAQSRRELFTLVAGLGVAFGSAKLFGVSGFALGAFAGMLLKESELSHKA  
ASDPLRDAFAVLFVSVGMFPDMILIAHPQVLAFTLTVTVGKSLAAPVIVRAFG  
HPTGTALITSLAQIGSEFSLAGLVQILAIIPETGRDLILAGALLSIYANPPLFSH  
LDKWAQAOADPATVPELPQGPPIQLDGHAIIVGYRGVSAQLLRGRGVPLVI  
DNDGSHVAKAAAGIPGIRGSAADRVLAEARPEQAKIATLAIPOPLEAGEALAKLA  
LNPSTLLAARHSDAEVKKHLEHGAADGVAERLAYSLSAEMVNSTPPYRALRPAS"  
BASE COUNT 1771 a 3401 c 3161 g 1785 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 0.000253 Length: 10118  
Score: 193.00 Matches: 94  
Percent Similarity: 35.52% Conservativeness: 47  
Best Local Similarity: 23.68% Mismatches: 144  
Query Match: 9.43% Indels: 112  
DB: 1 Gaps: 14  
US-09-889-746-2 (1-400) x AE012418 (1-10118)  
Qy 15 LeuSerValAlaValThrGlnGlnLeu-----TyrAlaGlnPro 27  
Db 3546 CTTTCGATTGCGGTTTCATCGCGTCCGCAACGGTTCGCGGTACCGCGCGCA 3605  
Qy 28 AsnGluSerLeuProThrValGluLeuLeuProValValIleThrIleAspLysSerGly 47  
Db 3606 CAGCAGCG 3653  
Qy 48 MetAlaLeuAlaAsnArg---IleThrGlnMetProHisThrThrLysValIleTyrGlu 66  
Db 3654 ACGCGCGTGGCG 3713

```

Qy 67 GIUGLIILGLINGLINATHRGYSERARGINLEUALAAPVALMICALINLEU 86
   ::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3714 CAGTCGTTG-----CAGGGACCGGGGACCGATATGCCACCGCGCTGGCAAGCTG 3767
Qy 87 ILEPDSERLEU-----GLYVALSERSERGLYTHRSHRANPHEGLY 101
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3768 CTGCGCTCCCTGCACTCCCGCGCGCGCATTCGACGATGACGCGCGCGCGC 3827
Qy 102 GLNTHMETHIEGLY-----ARG 107
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3828 GCCACGCTGGCGCGCGCTGCGCGCGATCCGCTGCTGCTGATGCGAAGCGCTAT 3887
Qy 108 GLNVALGHPHELEUENGLYVALPROLEUTHRGLYSERARGAPLIESERARGIN 127
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3888 CACACCTGCGCGCTGATCACTCAACCCGATATGTGGCGGGCGCTCGCGCGCGCGAC 3947
Qy 128 LEUENSERILEANPROANGLINVALAARGILEGLYVALLEU---SERGLYALATHR 146
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3948 CTCACCTCGCTGCGCGATGTCGCGCGCGCGCGCATCGAATGCTGCGTATGCGCTCG 4007
Qy 147 SERILETHRGLYSERGLYALATHRGYLYLEULLEANLLEVALTHRYSERAPLEU 166
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4008 GCGCAATACGCGCTCCGATGCGATGCGCGCGGTGTCACATCGTCTCAAG----- 4058
Qy 167 GLUGLUGLINPHEGLUHRARGILEGLYVALHISERYSERYSERGLUGLY 186
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4059 -----CACGCTCCAGACGCGCGCGCAACACG 4085
Qy 187 ILEGLYTRGLINVALGLYINSERVALALEGLYVALSERGLUANGLYANVALLEUALA 206
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4086 GTGTCGCTCAACGCGTGTTCATGACCAAGGCGCGACGAGCAAGCGCATCGCGC 4145
Qy 207 ARGLEUASPLAASPLARYZATHRTHRGLYLYALAPHEASPLAANGLYLYSARGILE 226
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4146 TCGGTGCGCATGAGCTTCGCGCGCGCGCGCATGACCGCGCGCTGCGCTG 4205
Qy 227 ALA----- 227
Db 4206 TCGTGAACCTACGAAACACATGAGACACACCGCGCGCAACACCGCGCGCAC 4265
Qy 228 -----PROGLUPROALAGLINTHRASPLASGLN-----ASPSELYS 239
Db 4266 ACCGTGCGCGCGACGCGCATTCAGAGTGGCGTCCGACGCGCGCACCGCGAG 4325
Qy 240 SERLEU----- 241
Db 4326 TCCACGTTCTACAGAGGCTGCGCGCGCTGCGCTACGATTCGCGCGAAGCTGGA 4385
Qy 242 -----SERVALASPLAASPLAASPLAASPLAASPLAASPLAASPLA 4445
Db 4386 TACGGCCACTTCAGTGCAGCGCGCGCGCGAGTACCTTCACAGCGCTATTCAC 4445
Qy 253 ASPPLASGLINLEANLEANLEUALALEUTHRHSIYANASPLASPLASPLASPLAS 272
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4446 GACACACCGCGCGCGCATGCGCTGATTCAGAACGCGCTTCGCGCGCGAGATTCAC 4505
Qy 273 ALAPROASPLARYZATHRTHRGLYLYALAPHEGLYLYLYSPROSERLEUASPLA 292
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4506 AACCGG---GTCAACGACCGCTGCGCGAGTCTC---GGCATCAAGGCGACAC 4559
Qy 293 ILELYSERLEUASPLASGLINPVALYTHRTHRYSERTHRPHASPLAASPLA 312
Db 4560 GAGTGAACCTGAGATCTGCGCGCTGACGCGCAAGACGATGAGCTTCACATCTTC 4619
Qy 313 TYRHSIYANASPLASPLASPLASPLASPLASPLASPLASPLASPLASPLAS 329
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 4620 -----AACACGATCAACGCAACCGCTCTAC 4646

```

```

ACCESSION  AB012012 AB008923
VERSION     AB012012.1 GI:21109993
KEYWORDS
SOURCE
ORGANISM   Xanthomonas axonopodis pv. citri str. 306.
            Xanthomonas axonopodis pv. citri str. 306
            Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
            Xanthomonas.
REFERENCE  1 (bases 1 to 10835)
            da Silva A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Purlan, L.R.,
            Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
            Jr., N.F., Alves, J.M.C., do Amaral, A.M., Bertolini, M.C.,
            Camargo, L.E.A., Camarotte, G., Camarvan, F., Cardozo, J.,
            Chambergro, F., Clapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,
            Curreino-Santos, J.R., El-Dorri, H., Faria, J.B., Ferreira, A.J.S.,
            Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,
            Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
            Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,
            Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,
            Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,
            Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,
            Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,
            Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos
            Santos, M., Truffi, D., Tsai, S.M., White, F.F., Secubal, J.C. and
            Kitajima, J.P.
            Comparison of the genomes of two Xanthomonas pathogens with
            differing host specificities
            Nature 417 (6887), 459-463 (2002)
TITLE      JOURNAL
            MEDLINE
            PUBMED
            12022145
            12024217
            2 (bases 1 to 10835)
            da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Purlan, L.R.,
            Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida
            Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C.,
            Camargo, L.E.A., Camarotte, G., Camarvan, F., Cardozo, J.,
            Chambergro, F., Clapina, L.P., Cicarelli, R.M.B., Coutinho, L.L.,
            Curreino-Santos, J.R., El-Dorri, H., Faria, J.B., Ferreira, A.J.S.,
            Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C.,
            Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite
            Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A.,
            Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J.,
            Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M.,
            Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A.,
            Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F.,
            Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos
            Santos, M., Truffi, D., Tsai, S.M., White, F.F., Secubal, J.C. and
            Kitajima, J.P.
            Direct Submission
            Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de
            São Paulo, Av. Prof. Lineu Prestes 748, São Paulo, SP 05508-900,
            Brazil
FEATURES
source     location/Qualifiers
            1..10835
            /organism="Xanthomonas axonopodis pv. citri str. 306"
            /strain="306"
            /db_xref="taxon:190486"
            /note="pathovar: citri"
            135..497
            /gene="pmps"
            /note="XAC3614"
            135..497
            /gene="pmps"
            /note="identified by sequence similarity; putative; ORF
            located using Blastx/Glimmer/Genemark"
            /codon_start=1
            /transl_table=11
            /product="6-pyruvoyl tetrahydrobiopterin synthase"
            /protein_id="AA038457.1"
            /db_xref="GI:21109994"
            /translation="MPNDIFRVFLTEAAHRLPNVPQGHKARLHGHSRVEIAYSGE
            PGATGWMIDFGDIAAFQPIYDRLDHYHNEIEGLNPTSRILAIWIWQDKPALPO
            LSEIVHETGSCRYRG"
            691..1026
            gene

```



Db 6223 ---CAATCGCTGGACACCGTCACGTC----- 6200  
 Qy 48 MetAlaLeuAlaSerArgIleThrGlnMetProHieThrIleValIleTyrGlu 67  
 Db 6199 ---ACTGCACGCGAGATTGGCCGAGCCCTGGTCTTCGTCATCACCAGCGAA 6146  
 Qy 68 GlnIleGlnGlnAlaThrGlySerArgGlnLeuAlaSerValMetAlaGlnLeu 87  
 Db 6145 GATATCGAAGCG-----CGCCCGCGGTCAACGATCTTCGACGCTGCTG 6101  
 Qy 88 ProSerLeu---GlyValSerSerGlyThrThrSerAsnHieGlnThrMetHieGly 106  
 Db 6100 CGCAGCATGCCCGCGTCATCTCACCGACAGTGCCTCCGCGCATGAGCGCAATAC 6041  
 Qy 107 ArgGlnValGln-----PheLeuLeuAsnGlyVal 116  
 Db 6040 CGCCAGATCATCTGCGCGCATGCGCGCGGAAACACGCTGATCTCGTCGATGGCAG 5981  
 Qy 117 ProLeuThrGlySerArgAspIleSerArg----- 126  
 Db 5980 CCGGTG---GCATCGCGCGATGCGGTGCGCATGGCGCGGTAGTGGCGAGCAATACAGC 5924  
 Qy 127 ---GlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGlnValLeuSerGly--- 144  
 Db 5923 GCGGACACCACTGGGTGCGCGCGCATGCGGTGAGCGCATCGAGTGTGCGCGCGCG 5864  
 Qy 145 AlaThrSerIleTyrGlySerGlyAlaThrGlyLeuIleAsnIleValThrIle--- 163  
 Db 5863 GCGGACGCGCGGTACGCTGGGCGCTCCGGTGGCTGTGTCATCATCATCACCAAGCA 5804  
 Qy 164 ---SerAspLeuGlnGlnGlnGlnPheGlnThrArgIleGlyVal-----His 178  
 Db 5803 CCCACCGCGCATCTGGCGCGTGCAGT-----ACGCTGTTCGCTGTCGCGCGACAC 5750  
 Qy 179 GlySerIleLeuSerSerGlyGlyIleGlyTyrGlnValGlnSerValAlaGly--- 197  
 Db 5749 ---AGCGCCGAAAGCGGTGTCACGCGCACCGGATTCAGTTCAGTGGCGCG 5702  
 Qy 198 ValSerGlnAsn-----GlyAsnValLeuAlaArgLeuAspValAsp 211  
 Db 5701 CTGTCCGACACGTTTGTTCCTGCTGTACGCGCAAT---CTCAACAAAGACCGATCCGAT 5645  
 Qy 212 TyrArgThrThrGlyAlaPheAspAlaAsnGlyLysArgIleAlaProGluProAla 231  
 Db 5644 TCCTGAGACTCATGCGCGCTTTCGAGTAAAGCGCGCGAGCGCGCG----- 5594  
 Qy 232 GlnThrAspIleGlnAspSerIleSerLeuSerValAsnThrAsnValAspTyrGlnLeu 251  
 Db 5593 ---GCCGAGCGCGAGCGGTGCGCAACAGACATCAATGCGCTGCGCTGGATCCG 5537  
 Qy 252 AspAspIleGlnAsnIleAsnIleAlaLeuThrHieTyrAsnAspIleAspThrAsp 271  
 Db 5536 ATGGAAGGGCAAGTGAATCGAATCGATCGCGCTTC-----AGCGCGCAAGCGCATCTG 5483  
 Qy 272 TyrAlaProAsp-----TyrGlyAsnArgLeuAlaValLeu 283  
 Db 5482 TATGCGGCGCATGCGCGCGTCAAGCGCGATGTCGCGCGCGAGATCTCAGCTGTTG 5423  
 Qy 284 PheGlyGlnIleProSerLeuAsnAlaIleLeuGlyLeuSerLeuSerGlnIleProLys 303  
 Db 5422 TTCGGC-----AGCGAAACCAACACACC 5402  
 Qy 304 ThrThrIleSerThrPheAsnIleAsnTyrHieIleAspAspLeuTyrP-----GlyAsn 321  
 Db 5401 ATGTTCGCGCGCGATCTCTGCTCAC-----CATGCGCGCACCTGGCGCGCTGGGAACC 5348  
 Qy 322 ThrIleAsnThrAsnAlaTyrTyrArgArgIleGlyIleArgPheTyrProPheValAla 341  
 Db 5347 TCGGCGCTCACCGCTGCTCAAGAGCAACGACCAACCGCGCTCAACGAAGCGCTGGCC 5288  
 Qy 342 ---ProPheSerIleAlaLysAlaLeu-----ProIleLeuGlnSer 354  
 Db 5287 GGTGGCCGGAAGGCTTCATCGCCACCAACCTGCTCGATTCCACCGCCACGCTCGATAAC 5228

Qy 355 MetAsnLeuProSerAla---ThrLeuAspAlaTyrThrIleAlaProGlnAlaArgAla 373  
 Db 5227 CTCAACATGACGCGCGAGTTCACCTGCACATGCGAGCATGGCGCGCAAGAGTCTGACA 5168  
 Qy 374 TyrGlyValLeuGlnSerGlnSerLys 382  
 Db 5167 TTCGTTTCGAGCGCGCCAGAGCCGT 5141  
 RESULT 42  
 AY008342/c 10399 bp DNA linear BCT 23-APR-2001  
 LOCUS Pleiomonas shigelloides heme iron utilization locus, complete  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Pleiomonas shigelloides.  
 Pleiomonas shigelloides  
 Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 Pleiomonas.  
 1 (bases 1 to 10399)  
 Henderson, D.P., Wyckoff, E.E., Rashidi, C.E., Verlei, H. and  
 Oldham, A.L.  
 Characterization of the Pleiomonas shigelloides genes encoding the  
 heme iron utilization system  
 J. Bacteriol. 183 (9), 2715-2723 (2001)  
 MEDLINE  
 PUBMED  
 11292789  
 2 (bases 1 to 10399)  
 Henderson, D.P., Wyckoff, E.E. and Payne, S.M.  
 Direct Submission  
 TITLE  
 JOURNAL  
 TEXTS of the Permian Basin, E. University Blvd, Odessa, TX 79762, USA  
 FEATURES  
 source  
 1. 10399  
 /organism="Pleiomonas shigelloides"  
 /db\_xref="taxon:703"  
 complement(186..943)  
 /gene="hugZ"  
 complement(186..943)  
 /gene="hugZ"  
 /note="heme utilization protein"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hugZ"  
 /protein\_id="ANG23392.1"  
 /db\_xref="GI:13774059"  
 /translation="MKEDIVMTNREQBELONRLOPEIAFRNRECRVLVATVDADGMP  
 NVSVAPVLTQDGYVVLSEIYARHARNQVPRSTIAMIEDESSQROLPAERLTPDA  
 RVEKARDTASQOEVALAEARHGDIVNLSOMADPILFRNVEGSLYKRGGAFOV  
 SSDDLVDVHLQEGHRTVSGEPVXA"  
 complement(1108..1590)  
 /gene="hugX"  
 complement(1108..1590)  
 /gene="hugX"  
 /note="heme utilization protein"  
 /codon\_start=1  
 /transl\_table=11  
 /product="hugX"  
 /protein\_id="ANG23393.1"  
 /db\_xref="GI:13774060"  
 /translation="MSASTQELARFVSIVQLVAALPEANAVMAHGSMAOSLBEET  
 DMGPLTTIVESGSLFERKGAEPVGRKPHGYFNLYSKGTGLHGLILSOVSHIALLA  
 KPFMGKASYSLQFPANGRITWFKVLYGNKQRLPEQVALFEQLRRLIAQPPDPAE  
 "  
 complement(1814..3169)  
 /gene="hugW"  
 complement(1814..3169)  
 /gene="hugW"  
 /note="heme utilization protein"  
 /codon\_start=1

```

/transl_table=11
/product="Exbb"
/protein_id="AAG23394.1"
/db_xref="GI:13774061"
/translation="MMMDLSVLDSDSILGHHSPPDLAVAFYTKRQAHAGSGRPLPPP
AVQGLQOLLQMDTPAIRPCRLIHIPECRVCTFCSPFOVAASPTLVAEYFAALSEE
LKRKAMPTQAAPAAAYIIGGPTDLTAQIIFALGQLIKAHPLPLTDDCHITLEGRL
NRFDEKTHAASVGGVNRFSFGVSGFDVRRKARLDPREKILLETVSHLVALDRAPI
IVDMFGIPLSATAMQODLQDFLHSGAGHVDLYOLIEMOQGTMPARLVEENRLOAPAA
PEKSRMYAQGVAFMQHGIQOIISVNHWTDRTRERSYNHLAKTTADVLPAGAGGPF
IDGALMOTRDLSEVHOAI REERYPVMMMSPNQHALHCAIKAGFDRGRLOREBELS
LSGHALYDVSPFLLTQWQERGSLLEADQARLT"TAGSPWNINMAQGLINALRLSPPLA
A"
complement(3169..5235)
/gene="huga"
/complement(3169..5235)
/gene="huga"
/feature="heme receptor"
/codon_start=1
/transl_table=11
/product="Huga"
/protein_id="AAG23395.1"
/db_xref="GI:13774062"
/translation="MPVQPSAIPRSTVSATAAMTAKHTGPKSALSLILNSVASF
AANAPAPFQQLNEVLVTARENKALSOETRSVAVVTOAQIQQQSPASVAEALKYIP
NVDVSGPRANAQSPSVRGLSGNRILQVVDGVRQNTTSGHVLRTSDPEMLSSLEVIK
GPTSLWSGALGGVAAQOTISAPDMLHGDQSGVGLKQGYASANDETKTSGAVYGLL
GDSVDFLNGYADANNLNGNHLTDSAYRNQGLAKFGWQADDAQRLSLSVRQAE
THONAPSQSVSTVPLVOOKTRDNTTLDYRLNPTDSRLWDARVYVYNKTEFDE
YRVTKQSDKYDRTGLNINSHSQFDLLSLTYGGDIYEDKTOGERGSKNRPADGR
SKWGSYQVADIPGQSNLLPLGKIDHTAEDKNXNIGSERSEDLSPSVGRWAATD
WLTNARYDEAFRAPSMEEMTYTGTFCWGMGNVFNKPNLKPETAKNKEISAQMR
FNNVLADDEAFATGYFHNYSNTIDQVNMFTQYVNVNTDARLGRVELDARYNRDL
ETLSYAOCTEGCKTNQPLNNIPAHKVLGVSHVFMDRDLKAGVNVSHYSQBELPS
NVTNVPYTYLVOLYTTWQPTGALKAVKVDVGIDNVTDVEYRQAFDQLYSAGRNFKY
GVRYSF"
5428..6294
/gene="tonB"
5428..6294
/gene="tonB"
/feature="heme transport protein"
/codon_start=1
/transl_table=11
/product="TonB"
/protein_id="AAG23396.1"
/db_xref="GI:13774063"
/translation="MPLTLLTAGAHLRNRKLLLSGLSCMSAHAALLFLGLQPTOTRA
VTLQSLSGPALVNLNTATGAQSVNTNAAADVQAPAKTEPKKVEPSAVPAKAA
ETAKTQAQKQAKPVLKKEPAPKVPAPKPEEKAPKTKVTKERKSTDNTEQNPVT
ANTGTAHSSATTSASGASNTASHAAPASTAQAGSSSSSEPRFRVPPSPAPEYKASR
MRQGTVLLEVKLTQGEQLQVLLKSSGFPFLDRSALKAVKQWQFLPQBIQGGQVS
HVVRIPVRFEIT"
6361..7170
/gene="exbB"
6361..7170
/gene="exbB"
/feature="accessory protein"
/codon_start=1
/transl_table=11
/product="ExbB"
/protein_id="AAG23397.1"
/db_xref="GI:13774064"
/translation="MILAQLHQLGSMWMLPILTSFLTGVILLDRFLRLQAGRLGQH
HLAQLPCHSNTNLSVASOLDTMOQLQORSLLLQGVALLLTHRNENKALREETA
AIWQSKRROLSSGKLNLNTAATSPILGLLTGLIQLQWQDQIGASNPVTPALLAD
GLGVAMTYTAGLFTALPQSLFSLMIDRLINGAEHGMNLYMLEGRICVRIS
HDPRLSAREQWDRSHPAAHHLYRTGVFIADRCQAYSEHGHLAGSGR"
7016..7447
/gene="exbD"
7016..7447
/gene="exbD"
/feature="accessory protein"
/codon_start=1
/transl_table=11
/product="ExbD"
/protein_id="AAG23398.1"
/db_xref="GI:13774065"
/translation="MIRDSAPSSSGLDITPLLDIIFIVLVFLLLLTANVRIOQSMDI
TLPEADSTOACTVPESLTLHLFADAPYALWGEOKFTDNAPSNALLTARAKPTSG
AQBEIIAADRDASTERLLNLLFLQRQNIIPAAQILTKGSE"
7444..8391
/gene="hugB"
7444..8391
/feature="periplasmic binding protein"
/codon_start=1
/transl_table=11
/product="HugB"
/protein_id="AAK38770.1"
/db_xref="GI:13774066"
/translation="MKQMLSRVALLACLSPVTLHAAPTDSAANTHVAASAIPP
TGQSTPKIAPTRVIGAGSNVTETIIVALGAGDVLGVDSRSSRLPASPAAATLGTIROL
PHEGMALRPDYVIGDEMAPASALTALQOAKIPLTVLPSTTPEALKAQVTOIAALL
HKEAAGQTLQTKIDQOLEQATRAQLQGHGKPLKAVYLLLDHGRPPMIAGNTPA
DTLLTLAGAVNPAAVRNQYVSAESLLAMQPOLIVVSRNMTGDADSLAKQPLLNQ
TPAGRNKAVLPINGKALIGGIGLSTLNEAQRANHLNQ"
8404..9465
/gene="hugC"
8404..9465
/feature="inner membrane permease"
/codon_start=1
/transl_table=11
/product="HugC"
/protein_id="AAK38771.1"
/db_xref="GI:13774067"
/translation="MLRLTLQSGSSGLLLSPMLLSIATGPVRLPLSLSQALLSL
PSHANDLSTOQLIOTIRLPTLLCTISVGGILALAGAVLOGLFERNPLADPGLIGVSG
AALGAGSIVLLPAGVSLAGSLAGSLGSLWGLYALCALFCGLMTLLVLYRIGRTP
FGTSVNMLLAGVAIGAIFSGGLLSYLADDDQDLRLSUQMSLAGAWDNTLLLS
TLVLVLVGFYRQAPALNALLGSEARHLGVNVRKRLIIGCAVIGIGVSVSGM
IGFGLIVPHLVMLCGPDYRNLLPLSALCGAILLLADLLARQLAAPAIPVGIITA
LLGAPFIFILLIQKRSAA"
9476..10264
/gene="hugD"
9476..10264
/feature="component of the inner membrane permease"
/codon_start=1
/transl_table=11
/product="HugD"
/protein_id="AAK38772.1"
/db_xref="GI:13774068"
/translation="MRNLTQGRRRQILSQLSLDRAGELTVLLGPNGTGKSTLLKCI
SGMAAQEIRLRFQPOSOWPILARRMAILPQSSALSFLAREVVALGRPLPHASG
KIADVEIQRCLQAVGAELHADSAYTVLSGEKORIHFAVLAQLDHEHMDVSPOTERE
PTAAAPKLLMDSEPTSLDLSHQHOTIQTQARRARQGDVILVILHDLNLAARYADRL
LLMHGQIQAGDSFEELVTAERIKQIFAFDAQVFRHPETGRHLHS"
2347 a 2933 c 2718 g 2401 t
BASE COUNT
ORIGIN

Alignment Scores:
Pred. No.: 0.000475 Length: 10399
Score: 189.50 Matches: 79
Percent Similarity: 47.65% Conservative: 53
Best Local Similarity: 28.52% Mismatches: 96
Query Match: 9.26% Indels: 49
DB: 1 Gaps: 17

US-09-889-746-2 (1-400) x AY008342 (1-10399)
Qy 10 LeuSerLeuProLeuLeu---SerValalaValThrGlnGlnLeuTyAlaGlnProAsn 28
Db 5142 CTGAGTCGTGTTAAATCTTGAACAGCGTGGCGGCTTTCATTTGCGCCACGCCACCGCT 5083
Qy 29 GluSerLeuProThrValGluLeuGluProValValIleThrIleAspLeuSerGlyMet 48

```



```

Db 5082 GAG-----CCTGAGACTGAGCTCAATGAAAGTCTGTTCAGCGCAACCCCTGAAAAATAAA 5029
Oy 49 AAlaLeuAlaAenAryIleThrGlnMetProHISThrThyValIleTyrgluGluGln 68
Db 5028 GGCCTG-----TCGCAAGAAACCGCTCGTTCGGTGTGAACCGCAGCA 4981
Oy 69 ILeGlnGlnGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIlePro 88
Db 4980 ATTCAGAACACACACCGCCGACG-----GTGGCTTAACACACTG---AAATACATTCTCT 4930
Oy 89 SerLeuGlyValSerSerGlyThrThrSerAspMetGly-----GlnThrMet 104
Db 4929 AACGTGATGTACAGTGGCGGTCTCCGCGCAATATCCGCAAAAGCCGTGATGACGGGGTTTA 4870
Oy 105 HIGlYArgGlnValGlnPheLeuLeuAsnGlyVal---ProLeuThrGlySerArgAsp 123
Db 4869 AGTGGCAACCGCATTTTGGACGGTGTGGATGGCGTGGCTGACAGACACTTACCTCGGACAC 4810
Oy 124 ILeSerArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSer 143
Db 4809 GTG-----CGACTTACTTCATGGGATCTGAATGCTGTCAATCCATCGAAGTGAATCAA 4756
Oy 144 GYAlaAthr---SerIleTyrglySerGlyAlaThrGlyLeuIle-----AsnIle 160
Db 4755 GGGCCGACACAGCTCCCTGTGGGGGACAGTGGCGCACTGGGTGGCTGTGGCGCAAACT 4696
Oy 161 ValThrIleSerAspLeuGlnGlnGlnPheGlnThrArgIleGlyValHisGlySer 180
Db 4695 ATTTGAGCTCCCAATATG-----GTCGACGGGT-- 4669
Oy 181 LysLeuSerSerGlyGlyIleGlyTyrglnValGlyGlnSerValAlaGlyValSerGlu 200
Db 4668 -----GATCAGAGGTGTGGGGGCTATCTGAACAAGTTACCGCACCGCCAAATGAT 4618
Oy 201 -----AsnGlyAsnValLeuAlaArgLeu-----AspValAspTyraArgThr 214
Db 4617 GAAACCAAAACCAAGCGGTGCGTATGATGCGCTGTGGGGATTCCTCGATTCCTGCTA 4558
Oy 215 ThrGlyGlyAlaPheAspAlaAsnGlyLysArgIleAlaProGluProAlaGlnThrAsp 234
Db 4557 AACGGCTACTACCGCCGATGCGCAATACCTGCGTTGGGCAATGGCGACACTGACACTGAC 4498
Oy 235 Lys-----GlnAspSerIleSerLeuSerValAsnThrAsnValAspTyrGlnLeuAsp 252
Db 4497 TCGGCCTATGCAATACGGCGCGGTAGCC-----AAATTGGCTGCGACGGCAT 4447
Oy 253 AspIleGlnAsnIleAsnLeuAlaLeu-----ThrHisTyraAsn 265
Db 4446 GACGCGCAGCGTCTGAGAGTGTGCGCAAGCCGAAACCCATCAGAAC 4396

RESULT 43
AB011991 10153 bp DNA linear BCT 29-MAY-2002
LOCUS Xanthomonas axonopodis pv. citri str. 306, section 369 of 469 of
DEFINITION the complete genome.
ACCESSION AB011991 AE008923
VERSION AB011991.1 GI:21109804
KEYWORDS
ORGANISM Xanthomonas axonopodis pv. citri str. 306.
SOURCE Xanthomonas axonopodis pv. citri str. 306.
Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
Xanthomonas.
REFERENCE 1 (bases 1 to 10153)
AUTHORS da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Purlan,L.R.,
Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
Machida,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
Okura,Y.K., Oliveira,M.C., Oliveira,J.R., Pereira Jr.,H.A.,
Rosati,A., Sana,J.A.D., Silva,C., de Souza,R.F., Spindola,L.A.F.,
Taktai,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
Kiteajima,J.P.
Comparison of the genomes of two Xanthomonas pathogens with
differing host specificities
Nature 417 (6687), 459-463 (2002)
22022145
12024217
2 (bases 1 to 10153)
da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Purlan,L.R.,
Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida
Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C.,
Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J.,
Chambergo,F., Ciapina,L.P., Cicarelli,R.M.B., Coutinho,L.L.,
Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S.,
Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C.,
Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite
Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A.,
Machida,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Meidanis,J.,
Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M.,
Okura,Y.K., Oliveira,M.C., Oliveira,J.R., Pereira Jr.,H.A.,
Rosati,A., Sana,J.A.D., Silva,C., de Souza,R.F., Spindola,L.A.F.,
Taktai,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos
Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and
Kiteajima,J.P.
Direct Submission
Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de
Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900,
Brazil

FEATURES
source
location/Qualifiers
1..10153
/organism="Xanthomonas axonopodis pv. citri str. 306"
/strain="306"
/db_xref="taxon:190486"
/note="pathovar: citri"
complement(86..2440)
/gene="XAC3446"
complement(86..2440)
/note="putative; ORF located using Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/db_xref="GI:21109805"
/db_xref="GI:21109805"
/translation="MPSGRLLRVGCVNVVASREVPAFGAKRPLRLAPKSLAVLTTLAR
OGGVTRREOLLAERVPTLPTNVQVAVQVQLKRAFPASHGSGRSEHLETKGVRL
MAAVVEMAHAPASATLPEDAPDSDFEPAOHAGVATSPAGSDTYAPPIRARQCPVA
PVAAPPAATPAAPRORRSMALAVACVALLVVLGLAATVSKRAAPAPAAQAPVLS
PKRPQVITAGGFDLPSTSPDGMAVAVASLTGDRATSLTKVDIAVPRVLETGP
PGVSRLLPWSMPTGROIAFSPROAPGSCRMIAATGPAASRELVRCGAOMLSFSMS
PDGRALVFSMTGTGAVRMRLDIAETOMQALDYPAGQPDVAPRYSPOCNVYFL
RNPOGLGIMRTIPAEGVAAERLTHNDALIRGSMWLPDGSGLTIGRRVDEARLYRDLR
DHSLDVGVDAQAPVAHGHVYQKRPQGVGVYIDTINAGTIRRRSPSSRD
QPMIAPDGRQULIPASNRSGAYGLMWGDVSKRSVAILGLRPSQADWSQADSVRL
VSGDAAGHAALFEVTPSGQGVRLPVQPRPQAIHLPDPRLLVTLAAADGHTFAT
LYDRSSPMRALATLDQVSQLRDLRVSQVLTLSRSGGVQVQVAPSLADTLHADPQ
QBSRRVYSMAQGVNGOLHYLSSTDSQCAAYSELRSGGOCIDAKFTTINGFSLNH
DGLTVLSLAERDGSALAYMALPQHSVAVAVGVIANLSTLRKITS"
complement(2864..4741)
/gene="chic"
/note="XAC3447"
complement(2864..4741)
/gene="chic"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="thiamine biosynthesis protein"

```

```

/protein_id="AM38290.1"
/db_xref="GI:21109806"
/translation="MNAAPTFLVQQOQSLSEAVTQPIPGSRKIFVRSRADLQVPMRE
IALTRTPTLFGEEENPPLSVDTSGPYTDPRAVDLAAGLAPLRAQMTAERGDTLALD
GLSSFGFRGHDARLDVRRPARLPRVARQGANVTOMHYVARRGILTPENYVAIRE
NORLEAVTASLRKHQPCQAFGASIQORIITPEVREETARGRAILLPNNINPESEPMI
IGRNPLTKINANIGNSAVSSGIAEVEKLVNSIRNGDVTMDLSTGKHIEHTRWIR
NSPVGITVPIQALRKVDGRAEELTWDFRDLIEAQEGVDYFTIHAGVLLRVPIL
TAKRVTGIVSRGSLIMAKCLAHKKNFLYTHFEDIQIMKAYDVAFLSGDLRPGCI
ADANDAQFGLTGLTGLTGLTGLTGLTGLTGLTGLTGLTGLTGLTGLTGLTGLTGLT
YDLPGLTGLTGLTGLTGLTGLTGLTGLTGLTGLTGLTGLTGLTGLTGLTGLTGLT
KIAHAADLAKHGPCAQRVDNALSKARPEPRWDQFHLGLDPEKAKEPHDITLKDPAH
KLAHPCSCGPHFCMKITQDVRDYANEHGVSEAGLSTGMEKSAQFVAQAQVYRA
T"

```

## gene

5335..8133

/gene="btub"

/note="XAC3448"

## CDS

5335..8133

/gene="btub"

/note="Identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"

/codon\_start=1

/transl\_table=1

/product="TubB-dependent receptor"

/protein\_id="AM38291.1"

```

/db_xref="GI:21109807"
/translation="MTMTTLTPTMRRGSLAIAISLCVSTSAIAQEQTEKNTNLDKIE
VTGSRISVDVENSOPVLVLRQDIEKQLTSADVLRKIPSNAGATNTNTVNGGDGS
ATVSNLNIITRNPGAEVNAVUGXGGDGGORTSFDATGATSRGNLVLGLSVKED
IAGVMDREISAPVYGAASSAYSSYSGRDRANGTSAGTVLPDSATASDGRYPYR
QADQYFANTSEYGNAYKDNLYLTPQRDALYVQGNVALTDNVRFPDGLYNRRQSS
COLGAFPLPQADRFQSDVGFPAHNOTDQETQYDQVGNLNLNAGLSTGVTGLTAC
VYTLGEGFQADRFQSDVGFPAHNOTDQETQYDQVGNLNLNAGLSTGVTGLTAC
LDGGVAFQGVNPLPAGVTOELDYILFTAHNTYQVKNSSFTANLTDGLFELP
AGWMSFAGVHERESGSRDALLISAGTSGNSFTPTSGSYTLDYVTEVAIPLKED
VFGQLLSAAVRVDYNTSTGTSNRSFGFKWPKFVADLLIRGNATVGFAPRSIDELY
LGASDFDTYDPCSTNYSYSAIAATCAARGVPNAVYABNSAGNSGQTIIPEDY
TSNALQPEKNTTLGVFPSPASVQGLDSDVDMWKIEITDAITFSAQILSCQYF
NSAPCDLVTDATGAIQVLPVNVGEMQMGYDFARYKLPQTAIFGQPLMLTDSY
YVAVNEOKDDNSAWESYNGIYHTNNPNWRTGTLTLDMSYGNLAANWAVYYSKMD
YDSELDGNGYRHHVATAFNDMQVSYQLPWNATVRLGLNNVPRDRPPVALSANSFD
SAYAIQGRYSLOYTORF"

```

## gene

complement(8392..9939)

/gene="tar"

/note="XAC3449"

complement(8392..9939)

/gene="tar"

/note="Identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"

/codon\_start=1

/transl\_table=1

/product="chemotaxis protein"

/protein\_id="AM38292.1"

/db\_xref="GI:21109808"

```

/translation="MLSLPHVETAILREATRLADGSRPVPSIPLLQPLVARLQAEHE
DRTVLRTAQALAGLELLPGLAETEMGRRLQAGFNLAETIRQVALSVQITAG
VPLVAVDENVMVROQAEALDGAARVRESLQVQNAELAQVIALAASADDSAR
QGVAAHALGOAMHEVERAARAGEVIEIDSVAFQTNILSINASIEAAGDAGRF
AVATEIRRLERAAAARDVRAILLAETSAAVGEASASARTEAVLDITRVLQASS
AMTVAGRIHQDGIIRGERAVDGVVALGRSLNQHAAHVAERSEALRGTEETUDCV
GLFRLLPFDQVPRHALVQAVATSGKIGSALAQIARGQIQNALPARTYTIAGV
EPKPFSTFDALCDQLLPALQEPLEAHPIVFAICANPDGVYVPHNLRFTQPTGDA
KRDLVGNRTKRKFTDRVGRSVCAHTDPYRLQVYRRDTQGIIMFDLSTPIFVGGKHWGL
RVGYTLE"

```

BASE COUNT 1894 a 3413 c 3083 g 1763 t

## ORIGIN

## Alignment Scores:

Pred. No.:	0.000501	Length:	10153
Score:	189.00	Matches:	93
Percent Similarity:	38.3%	Conservative:	60
Best Local Similarity:	23.3%	Mismatches:	156

```

Query Match: 9.23% Indels: 90
DB: 1 Gaps: 17
US-09-889-746-2 (1-400) x AB0111991 (1-10153)
QY 15 LeuSerValAlaValThrGlnGlnLeuTyrAlaGlnProAenGluSerLeuProThrVal 34
Db 5389 ATCAGCCTGTCGTGAGCAGCGCCATCCGCCAGGAAACAGACCGAGAAATCCGACC 5448
QY 35 GluLeuGluProValValIleThrIleAspLysSerGlyMetAlaLeuAlaAsnArgile 54
Db 5449 AACCTGGACAGATCGAAGTCACC-----GGTTCGCGCATCCGCTCGTGCAT 5496
QY 55 ThrGlnMetProHisThrThrLysValIleTyrGluGluGlnIleGlnGluAlaThr 74
Db 5497 GTTGAATAATTCGAACCGCTGCTGCTACTCACGGCCAGGACATCGAAACGAG----- 5550
QY 75 GlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSerSer 94
Db 5551 GGCCTGACCTCGGTGCGGATGTCTCTCAAGCGC---ATCCCTCCAAACGGTGCACGATG 5607
QY 95 GlyThrThrSerAsnPheGly-----GlnThrMetHis 105
Db 5608 AACACCACCGTCAACACGCGCGCGCTCGGCCAGCGTGAGCCTCGCGCAACCTGGCG 5667
QY 106 GlyArgGlnValGlnPheLeuLeuAenGly-----ValProLeuThrGlySer 121
Db 5668 GCCTCAGCACCGCTGCTGCTCAATGCGCGGGTGGGTCTCGGGCTGTCCGGCTCG 5727
QY 122 ArgAspIleSerArgGlnLeuAsnSerIleAsnProAenGlnValAlaArgIleGluVal 141
Db 5728 GTGGAC-----CTGAACACCAATCTCCCTCGGCCATCGTGGAGCGCGTGGAAATC 5775
QY 142 Leu---SerGlyAlaThrSerIleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIle 160
Db 5776 CTCAGGATGCGGCTGCTGCTCATTTATGGCTCCGATCGCATCGCGGGGTGGTCAACATC 5835
QY 161 ValThrLysSerAspLeuGluGluGluGlnPheGluThrArgIleGlyValHisGlySer 180
Db 5836 ATTACCCGACCAATTCGATGGGGCGGCGAAGCAACCGCTACGTTGGCAGTACGCGC--- 5892
QY 181 LysLeuSerSerGluGlyIleGlyTyrGlnValGlnValGlnSerValAlaGlyValSerGlu 200
Db 5893 -----CAGGCGCAGCGTCAAGCGACCTCTTCGACGCCACCGTGGCGGCCACCGCAG 5946
QY 201 AenGlyAenValLeuAlaArgLeuAspValAspTyrArgThrThrGlyGlyAlaPheAsp 220
Db 5947 CGCGGCATCTGGTCATCGGCTTGTCC-----ACGCTCAAGGAAGATGCGGTGATG 5997
QY 221 AlaAenGlyLysArgIleAlaProGluPro-----AlaGlnThr 233
Db 5998 GCGGCGCAGCGGAGATCTCGCGCGAGCGAGTGTACGCGCGGGGTCTGTCGCCCTATAGC 6057
QY 234 AspLysGlnAspSerLysSerLeuSerValAsn---ThrAenValAspTyrGlnLeu--- 251
Db 6058 AGCTACTCGGCAAGCGCGCGCTCCGTGCAATGGCCACCGCGCGCTCGACGGTGTG 6117
QY 252 -----AspAspLysGlnAenIleAenLeuAlaLeuThrHisTyr 264
Db 6118 CCCGACCGCTACCGCGAGCGGCGAGCAGAGGTATCCGCCCTAGCCCTGGACCATAC 6177
QY 265 AenAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeuPhe 284
Db 6178 ACCGCTTCAACACATCGGAATATGGCTTACAACCTACGCCAAGGACAAATTACCTGCTCAG 6237
QY 285 GlyGluLysProSerLeuAsnAlaIleLysGlyLeuSerLeuSerGluGlnProLysThr 304
Db 6238 CCACAGAAGCGCAGCGCTGTGATCGTGCAGGCG-----AAT 6273
QY 305 ThrLysSerThrPheAsnIleAsnTyrHisAspAspLeuTyrGlyAsnThrIleAen 324
Db 6274 TACGCACTCAGCGACACGCTGCGCTTCAAGTTGATGACATG----- 6315

```

Qy 325 ThrAsnAlaTyrTyrArgAGGluIyGlyArgPheTyrProPheValAlaProPheSer 344  
Db 6316 -----TATAACCGCCCGCCAGCTCGTCAG----- 6339  
Qy 345 IleAlaLysAlaLeuProIleLeuGlnSerMetAsnLeuProSerAlaThrLeu----- 362  
Db 6340 -----CAACTGCGCGGCTTTCCGCTCGCTCGCAACACC 6375  
Qy 363 -----AapAlaTyrThrLysAlaProGlnAlaArgAlaTyrGly 375  
Db 6376 GGCATCTGTTAGGCGCGACACTACTACAC---CCGACACACGCGCTACGCG 6429

RESULT 44  
LOCUS AE011961 10417 bp DNA linear BCT 29-MAY-2002  
DEFINITION Xanthomonas axonopodis pv. citri str. 306, section 339 of 469 of the complete genome.  
ACCESSION AE011961  
VERSION AE011961.1 GI:21109495  
KEYWORDS  
SOURCE Xanthomonas axonopodis pv. citri str. 306.  
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group; Xanthomonas.

REFERENCE 1 (bases 1 to 10417)  
da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida Jr., N.P., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C., Camargo, L.B.A., Camarotte, G., Cannavan, F., Cardoso, J., Chabrego, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L., Curiano-Santos, J.R., El-Dorfy, H., Faria, J.B., Ferreira, A.J.S., Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A., Madalira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J., Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M., Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.H.A., Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinoza, L.A.F., Taktla, M.A., Tamura, R.E., Teixeira, E.C., Tezara, R.I.D., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and Kitzjima, J.P.  
Comparison of the genomes of two Xanthomonas pathogens with differing host specificities  
Nature 417 (6887), 459-463 (2002)

TITLE 2 (bases 1 to 10417)  
JOURNAL 12024217  
MEDLINE 22022145  
PUBMED  
REFERENCE  
AUTHORS da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Quaggio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida Jr., N.P., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C., Camargo, L.B.A., Camarotte, G., Cannavan, F., Cardoso, J., Chabrego, F., Ciapina, L.P., Cicarelli, R.M.B., Coutinho, L.L., Curiano-Santos, J.R., El-Dorfy, H., Faria, J.B., Ferreira, A.J.S., Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A., Madalira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J., Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M., Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira, J.H.A., Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinoza, L.A.F., Taktla, M.A., Tamura, R.E., Teixeira, E.C., Tezara, R.I.D., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and Kitzjima, J.P.

TITLE Direct Submission  
JOURNAL Submitted (28-NOV-2001) Departamento de Biocimica, Universidade de Sao Paulo, Av. Prof. Lineu Freire 746, Sao Paulo, SP 05508-900, Brazil

FEATURES  
source  
1. 10417  
/organism="Xanthomonas axonopodis pv. citri str. 306"  
/strain="306"  
/db\_xref="taxon:190486"  
/note="pathovar: citri"

gene 122..2509  
/gene="bfea"  
/note="XAC3166"  
CDS 122..2509  
/gene="bfea"  
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=11  
/product="ferric enterobactin receptor"  
/protein\_id="AAM38010.1"  
/db\_xref="GI:21109496"  
/translation="MNTPVSAALAVVAALASPALAQAATAPANTPTVITGT  
LRSDRTVASOSPIDIIISBSLOATGTELETALAGAAAPSLNPFPSALDSALRPA  
OLRGLSPQVVLVNGKRHTSSLLNLNGTIGRGAADVNTPTVPAALAYVEVGA  
SAOVSADIAAGVNVIVLKGARKGSIOAGPFOYAGDGNMISGDTVAAGDGMGL  
HYAQLNODPTDRARAYAGAPSVSQPVGKARKITGPEVNAQVSNANTEPATPDSI  
TGAFATNSNDITSPAFRAPGSTTONLSIYFGFLPEIQYAKDSLVAGVSGT  
ASGMDVASVYNGNRIDFHTNLTNLSLGPSPSFPYDALLETTONLVNDVKGFD  
WGLAYPTVAFGAERENKKNQSPGASVFOAGTLAGAGGAFSGVAGQVARD  
VALYADLEADPTDFESAGLAGEYEDSFGSOAGKFAFYADTKIALRGVSGFR  
APSLAQQYFOSSTTFLLGNNPFEIRPPADSNVAVLGAEPDAFTLSVSGIVT  
OPTALYITVYAYQIDVDRIYLSNLGTGVRSLSESQIFGICNGKVFPTNAVDTRI  
KGVDSVYRMLAASVDLTAGIVSDEVSVANPAALSNGLSLERIDRERGR  
IEEGPRDKPLDINSWNLHWTLAGATRYKYSTRPADNSNDQTPGAKVVDVMSASY  
KVDWTLTLGADNVLDVDEYDENNFANSTSGFPYSLSPFGFNAYVYGRINRW"  
2974..3369  
/gene="XAC3167"  
2974..3369  
/gene="XAC3167"  
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AAM38011.1"  
/db\_xref="GI:21109497"  
/translation="MIDHLVVDLPASRRFYQAVLDVLGIPIVGSGDFPMDELF  
VSDRTVAQGLTGHFLAFOARDVAMQAFHQALVHGNDNGAPGLRPHRYHYAA  
FVLDPDGNTAVFHVGAQRSAADVHTP"  
3580..6030  
/gene="bfea"  
/note="XAC3168"  
3580..6030  
/gene="bfea"  
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"  
/codon\_start=1  
/transl\_table=11  
/product="ferric enterobactin receptor"  
/protein\_id="AAM38012.1"  
/db\_xref="GI:21109498"  
/translation="WDARBSAPMHPPTPTSCYRNCKTSPPLAVVALLSPASSTA  
SAOSQAPKLDLTLIVGTAVARTVASASPIDIISPOLSTGTTELATASRAI  
PSLNFPRALISDGAVAPADQIRGSPQVVLVNGKRYSTALINDTQYSGSSPA  
DLNTPFAVARIETLRGASAOVSADIAAGVNVIVLKGSEGS INGRYKYSAGD  
BOYSLDGAGSPGKTGVHLLAAGSHDQNRALPRGREGORYGDDIDOGALSN  
GEYSPTDLYTYSRMSVRSRLVLSNGYFRSGDNRNRPETPDGFLPTIVNSDVSN  
VGLKSTVEGLANTLDSTNYGNNLTIPVRSNLNLSLVSPTDPAHTLVTQVNLN  
ADFTKLDWGLAYPTTLAFGAENGKRNQSPGAAASANGIISBANSLIPGAVFP  
GPKLSDAGYVNRNHSVADLEADTEFLSAGLGRYKSGDFGTATAGSLRLYAF  
DKVAVRAASGFRAPSLQOONFOSIAYPTVAVNOPNTVVAIPETIGPTPNPAI  
ALGAEPKAEESKVGGLVQVPEVNTIYDAVRIIDDDIVLSNLTNSAVNPAI  
ANGPRTGGRVPTNAVDTYQGVDAVYTYWNTDGSVELTSGNTYKTEIKRADN  
PALMALIPSAVRIGRAETGRITBGTPLDKRFLUGTNSPGWMSFGTATRNGEESTG  
TNAGSDQYAAKWTLDLASYLKAWNTVGVGNVLEYPDQAQAGLSTRYLPSSA  
SPFGNGLVYANVYKW"  
6520..8772  
/gene="bfea"  
/note="XAC3169"  
6520..8772  
/gene="bfea"  
/note="identified by sequence similarity; putative; ORF

located using Blastx/Glimmer/Genemark"

```
/codon_start=1
/transl_table=11
/product="ferric enterobactin receptor"
/protein_id="AAM38013.1"
/db_xref="GI:21109499"
/transl_table="MAESSEPIDIISEESLOATGATVATGALGKLPSLNPRPAIAD
NDARPAATLGLSPDNVLVLVDKGRYHTSSLVNPNPVYGRSGAPADLNLSLPMASAIAR
IEVURDGAAGTSDAIGAVNVI LKHHGAAGSSVNVGIMDKDQAGNIGDSVG
LPGATGSDSAPGWRVLSWYNSQNSMTNRSNTNRTTAGAANPSPVYQRHGDDPA
RFOGLVAGNFYPSVLYGLHSVSREVTSNGYRAYDSDNRVQAVYVYQFLPOI
YNPINDRAVLGVKGTSDWHDVSVATYKNDMTNLSINTNLWYSGTSPNEN
SGPTEQGTINADPNKSLGAYPVNLAQVHRODRYEIVAGSPASYEPDPTIN
PETGSPYGGQVSGSLPSVAGKFRHSNAAAYALEADLTKLGGIAGRYENYSDA
GSTSKLSARAFYDFTSMRATISNGFRAPSLAQONYASVVTLPDQGLIOVGTYRT
SDPTAVQIGARFVPSKSTNYGIGSWQPTTNFTSLDVYQIRIWDQILYDQLOLAQ
PIGOVAGQVFFNGATSRTRGVWVNSYRADLGAGSDLSASANYNKKILELSNPN
FGRASQGLTSTPRTKYMLAADWTLDFGSLNGNATRYGAIKRISDPADGSDQDTYSA
RWLLNLAANYTQALTPYTGADNVNTQYPTKAQLTTGVDYDRASGLQYSSLSLPPFNGR
YWYGRVYTRF"
```

gene complement (8972..10165)

/gene="bioI"

/note="XAC3170"

complement (8972..10165)

/gene="bioI"

/note="identified by sequence similarity; putative; ORP"

located using Blastx/Glimmer/Genemark"

/codon\_start=1

/transl\_table=11

/product="cytochrome P-450 hydroxylase"

/protein\_id="AAM38014.1"

/db\_xref="GI:21109500"

```
/translation="MLLSDLATPQPHDPYPTAYRLREGLVQVADGRILMSGRYAVV
DRLLASDRVGRDYQLSVRLRYCEAAVHLPLFGMSRMFLNLPPLHQLRGLMTQAFG
AROMESREVASDIAAGLIDAFQANGHCDLLEAFPLPIAICRMULDIAADYATLS
HATSAKLVFQPMATBELQATSVAYDOLATYFHVIAQRESAGSDDLIARFIOAEDN
GRUSEEIVSNVILLFPAGHETSNMICALVALHRHPQQLRLLOETFGLLPNVLE
CMRYDSVQMATRALQDIFEIGVAVPRTGWLMLGAAHNDITLQFTDQVLDLRQOQ
GRALSGGGIHHCLGNRLALIEVEAALACLILARLPALREQLDLTSLMNDNRANLRGVDA
LLASW"
```

BASE COUNT 1890 a 3554 c 3175 g 1798 t

ORIGIN

Alignment Scores:

```
Pred. No.: 0.000564 Length: 10417
Score: 188.50 Matches: 114
Percent Similarity: 37.03% Conservative: 63
Best Local Similarity: 23.85% Mismatches: 161
Query Match: 9.21% Indels: 141
DB: 1 Gaps: 21
```

US-09-889-746-2 (1-400) x AE011961 (1-10417)

```
Qy 5 HisTyrPheGlnTrpLeuSerLeuProLeuLeuSerValAlaValThrGlnLeuTyr 24
Db 6376 CACAACCTCAACGCACTCTCGGTG---GCGATTTCAGTCGCACCTACCGTCTGCTCTTC 6432
Qy 25 -----AlaGlnProGlnGluSerLeuProThrValGluLeuGluProValValle 41
Db 6433 GCTGCCAGTGGCAGTCGCAATCGACCCAGCCGACAGAAACCTCGACACCTGATCGTC 6492
Qy 42 ThrIleAspLeuSerGlyMetAlaLeuAlaAsnArg---IleThrGlnMetProHisThr 60
Db 6493 ACC-----GGTACGCGCGTCAGCAGCCGCGACGCGTGGCCGCAATCCGAATCGCGC 6540
Qy 61 ThrLysValIleTyrGluGlnIleGlnGluAlaThrGlySerArgGlnLeuAla 80
Db 6541 ATCGACATCATCTCCGAGAGTGTGTG---CAGGCCACCGCGCGCCACCGATGTCGCC 6594
Qy 81 AspValMetAlaGlnLeuIleProSerLeu-----GlyValSerSerGly 95
Db 6595 ACCGCGCTCGGCAGAGTGTGTGCGTGTGTAACTTTCGCGCGCCGCCCATCGCCGACGCC 6654
```

```
Qy 96 ThrThrSerAsnPheGlyGlnThrMethHisGly----- 106
Db 6655 AACGAGCGCCGCGTCCGCGCCACGCTCGTGGGTGTGCGCGGACAAACGTCGTGTCGTG 6714
Qy 107 -----ArgGlnValGlnPheLeuLeuAenGlyValProLeuThrGlySer 121
Db 6715 GTCGACGCGCAACGGGTATCACACCTCTCGTGGTGAACCTACAACCCCTATGTCGGCGC 6774
Qy 122 ArgAspIleSerArgGlnLeuAenSerIleAenProAenGlnValAlaArgIleGluVal 141
Db 6775 GGCTCGGACCGCGCGCATCTCAACTCGTTCGGATGTGCGGATCGCGCGCATCGAGGTG 6834
Qy 142 Leu---SerGlyAlaThrSerIleTyrGlySerGlyAlaThrGlyGlyLeuLeuAsnIle 160
Db 6835 CTGCGCGACGCGCGTTCGGCAATACGGGTCCGATCGCGCGGTGTCGTCGAACATC 6894
Qy 161 ValThrLysSerAspLeuGluGluGlnPheGluThrArgIleGlyValHisGlySer 180
Db 6895 GTGCTCAG-----CACGGCGCC 6912
Qy 181 LysLeuSerSerGluGlyIleGlyTyrGlnValGlyGlnSerValAlaGlyValSerGlu 200
Db 6913 GTCGCGCGCAGCAACAGCATT-----TCGGTC 6939
Qy 201 AsnGlyAsnValLeuAlaArgLeuAsp-----ValAspTyrArgThrThr 215
Db 6940 AATGCGCGCATCATGACAGAGGCGACGCGCACAGAACGATATCGAC---GGCTCGGTG 6996
Qy 216 GlyGlyAlaPheAspAlaAsnGlyLysArgIleAlaProGluProAlaGlnThrAspLys 235
Db 6997 GCCTCGCGCTTCGGCGCCACAGCGCGATAGCGCGCTGCTGGTGGTGGTGGTGGTGG 7056
Qy 236 GlnAspSerLysSerLeuSerValAsnThrAsnValAspTrpGlnLeuAspLysGln 255
Db 7057 AATTACCAAGATTGATGGACCAATCGCAGC-----GAG 7092
Qy 256 AsnIleAsnLeuAlaLeuThr-----HisTyrAsn 265
Db 7093 AACACCAATCGCGTACCACTTCGCGCGCCCAATCCAGCGCGCTGCCCTTACCAG 7152
Qy 266 AspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeuPheGly 285
Db 7153 CGTCATGCGGATTCGGAT---GCCAGGTTCTACAGGCGCTGGTGGTGGTGGTGGTGGT 7209
Qy 286 GluLysProSerLeuAsnAlaIleAspGlyLeuSerLeuSerGluGlnProLysThrThr 305
Db 7210 TTCAGCGCCAGCGTCGAACCTGTATGGCCATCTGAGCGTCAGCGCGCGGAGTGAACCTCC 7269
Qy 305 ----- 305
Db 7270 AACGCGTACTACGTCGCTACGACAAACAGTACCGCAATGTGACGCGGTGTATCCCAAT 7329
Qy 306 -----Lys 306
Db 7330 GCCTTCCTCGCGCAGATCTACACCCGATCAACGACCGCTCGCGGTGCTCGCGGTCAAG 7389
Qy 307 SerThrPheAsnIleAsnTyrHisHisAsp-----AspLeuTrpGly- 320
Db 7390 GGCACACGGAAGCGATTGGCAGCTGTGGACGCTGTCCACCTACGCGCAAGACGACATG 7449
Qy 321 -----AsnThrIleAsnThrAsnAlaTyrTyr-ArgArgGluLysGlyArg 335
Db 7450 ACCTTCAACATCTCTCAACAGCATCAACACCAATCTGTACTGGACGACGCGGAGTTCGCGC 7509
Qy 335 g-----PheTyrProPheValAlaProPheSerIleAlaLysAlaLeuProIleLeuG 353
Db 7510 AGAACTTCAATTCGCGCGGTTTGAACACGCAAGCAGCGCTCAACGCGGACCTTCAAC 7569
Qy 353 nSerMetAsnLeuProSerAlaThrLeuAspAlaTyrThrLysAlaProGlnAlaArgAl 373
Db 7570 AAGT-----CCTTCGACTGGGTCTG-GCCTAT-----CCGGTCAATCTGCGC 7610
Qy 373 aTyrGlyValLeuGlnSerGluSerLysAlaGluValLeuGlyArgValPro 390
```

```

Db      7611  CTTGGGTGTAACACCGCCAGGACCGTACGAGATGTCGCCGCGCTGCCG 7662
RESULT 45
LOCUS   AP003592/c                      348250 bp    DNA      linear    BCT-28-NOV-2001
DEFINITION
AP003592 BP. PCC 7120 DNA, complete genome, section 12/19.
AP003592.1 GI:17132210
KEYWORDS
SOURCE  Nostoc sp. PCC 7120 DNA.
ORGANISM Nostoc sp. PCC 7120
Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
REFERENCE
1 Kaneko, T., Nakamura, Y., Wolk, C.P., Kuri, T., Sasamoto, S.,
  Matanabe, A., Iriuch, M., Ishikawa, A., Kawashima, K., Kimura, T.,
  Kishida, Y., Kohara, M., Matsumoto, M., Matsumoto, A., Muraki, A.,
  Nakazaki, N., Shimo, S., Sugimoto, M., Takazawa, M., Yamada, M.,
  Yasuda, M. and Yabata, S.
  Complete genomic sequence of the filamentous nitrogen-fixing
  cyanobacterium Anabaena sp. strain PCC 7120
  DNA Res. 8 (5), 205-213 (2001)
  2 (bases 1 to 348250)
  Kaneko, T.
  Direct Submission
  Submitted (02-MAY-2001) Takakazu Kaneko, Kazusa DNA Research
  Institute, The First Laboratory for Plant Gene Research, Yana
  153-3, Kisarazu, Chiba 292-0812, Japan
  (E-mail: kaneko@kazusa.or.jp,
  URL: http://www.kazusa.or.jp/cyanobase/,
  Tel: 81-438-3935 (ex.2338), Fax: 81-438-52-3934)
  Location/Qualifiers
    1. 348250
FEATURES
source
    /organism="Nostoc sp. PCC 7120"
    /db_xref="taxon:103690"
    /note="synonym: Anabaena sp. PCC7120"
    /gene="a113116"
    /complement(58..795)
    /gene="a113116"
    /complement(58..795)
    /note="ORF ID: a113116"
    /hypothetical protein"
    /codon_start=1
    /transl_table=11
    /protein_id="BAB74815.1"
    /db_xref="GI:17132211"
    /translation="MEPLSSIEDAIAQGREATKSLADGCRIOVELLPPELKMFPV
    AQFLPCLVEYSRLKVPFADAGAAALARDKDAFOLLIDGTGSIASIOKIPED
    BIFLIPSPVSEPOLKCEIIGDRPAIFLNPRLEADAGTGIYTAROTREPRINI
    GSCYLRPIIDDETALEFSYRGDMETIVENDGNWAKIANLPKKSDBEDILIMKQPO
    TGAADAVPARKSVKSLORFKALNS"
    /complement(1024..1287)
    /gene="a113117"
    /complement(1024..1287)
    /note="ORF ID: a113117"
    /hypothetical protein"
    /codon_start=1
    /transl_table=11
    /protein_id="BAB74816.1"
    /db_xref="GI:17132212"
    /translation="MMVESNRMDKNNFSAARSRYGRGAENLTENADLREANRIGY
    VTALETSGLSPDEACKOIKSLMKOLKRSKALISNDPMT"
    /complement(1635..2879)
    /gene="a113118"
    /complement(1635..2879)
    /note="ORF ID: a113118"
    /unknown protein"
    /codon_start=1
    /transl_table=11
    /protein_id="BAB74817.1"
    /db_xref="GI:17132213"
    /translation="MRSEIMAEOKSKNYWIRLNALIVSMSSLVVANYTLPKALDI
    PENSGLIAIKTPDEROEAVOQAEIOEIPSTLPATPOQKOSIESIPRESNST
    DSNTPPORENSAPRISTSSNPNASQPOPDINASPTSSNSRPTNSASAPORC
    OSITMPORGRNSTAKISPVSLNPFREINVDIAFGYLSKGDPSORRHFHYFEG
    RENOLQIRLGSTDSRSLNLSPPFLIPDPKVLTKGDSVSTNSIKDAFVFR
    LPVGTITIAVTSRNPSTGRYSIALRNDRASVYILDESAELNAAOSTIKORSAVNS
    OFQKKNQVLIRADISINEFSPYVILLNSQKTIADNDQDGLYSALIDBARLPEDD
    TYIVIVTSNPNNGTVRLTLF"
    /gene="a113119"
    /complement(3345..5375)
    /note="ORF ID: a113119"
    /product="WD repeat protein with Ser/Thr protein kinase motif"
    /codon_start=1
    /transl_table=11
    /protein_id="BAB74818.1"
    /db_xref="GI:17132214"
    /translation="MLCCINPNCSPMPQPDGMYCQRCNTOLIPLRGHYRIKVLSD
    BGGRTYLSIDIDIKNELCVKQEPAPVQVSHAMKKAVALFKQEAOLHIGHEHO
    PTLAIFRQDNYLPLVQPFINGNNILQELRGQVYNSTIYFELLDLPVLTHERG
    VTHRIDIKPQNTIRKQSDGLVLDIPGAQKQKATQTOGLTIGSLGTPLEQWYQYK
    AYPASDLEFSGATCFHLTGINPSNLFVEQYSWVESWQOYWNSSNSDNGEYLVKY
    LNKLETDIQRVQSADEVMNDLIKORSLRLKTTIPKSAIFSRSGASTSLASTT
    KQAWKILGRKQOLINTMSALLGLGVGHVGLPOLITKFSISTOPYTLKGAS
    DVNSAFSPNBEPLFASGSDDKTIYVNNKXOKIHTLPHSGGWMAIAFSPDGKTLAS
    TGADVTITLNNLAKGERIRHKGHSQVSAVAFSPDGKTLASGIDTKIKAMPATGK
    EIRLQHSRSGVAVNAFSPDGKTLASGSDMTIKLNNLTVSVKVIHTLGHSDLVGSDN
    FNSDQTLASGSKDKTIKLNLSGKTIKRLRGHSDKNSAVYAPRSDVLAWSGND
    TIKWNLTYGRIKRLKDSGYIVSIVSPGRNLASGSAENIIKIPMWS"
    /gene="a113120"
    /complement(5827..9687)
    /note="ORF ID: a113120"
    /codon_start=1
    /transl_table=11
    /product="two-component hybrid sensor and regulator"
    /protein_id="BAB74819.1"
    /db_xref="GI:17132215"
    /translation="MGILKNKRSFSFMNCTTSLPSHLKSAIRDPYIIRPETVMD
    AIAQMSGVACVDTAKLDEHLEARSQCVLIVAAGRLGIFTERVNVGLCSQRCLN
    LAIREVMHPVVTLESEFDTLPFAVNLQOYRIRHPIIDQERVVGVLINSESRHT
    SRSDIARLRIVSEVMCEVICATPDSKMLIAQOMTENRSSWITVQGLQFQPKPQ
    IPVGIYTERDVIQFQALGSLNETCQVQWMSPIFAVPDLSLRFVQIMQRLIRL
    AVTDQGLGIVTQSSLLQALNPLEVYKLAQMBEKVRLETERVQVQIMQRLIRL
    LVADRTASIKAKAGREQVSDISROISLSLQITLIDTVQVQVQLGCDNVNIFYE
    TEMESIVVAESTSTSLGRIEELCOOSABIRQGRIRVVDIYTIEMTDCHRE
    FLIRQTPAKTILVPLCGDELMGLINSESQHPREMOETEVELLOGSLVLAIAIHOA
    TTHQKQQLALAQORSQQLGSSORVITTSQNLAKIEETAEIQAREAOLOQTSDR
    LSLSKSGAICGMDIVENNI FMDERTYELGYATKSDSLVDTMANRLPDDRAN
    IESVVOQAILQGAEDYEFVYIHPDGSIHFKAYGVNADGEGPOXNIQVHPIFSR
    KESBEIRIQOYEREKLEIETORIRQSDLDQITIDICQVCOVLOADRVGIFKPYE
    SNFNDGEFVASVVEFSPVYAIRVHDEFCENASTYALGSRVYVDDIYHSDMTT
    TDIIAOFHVRANVIMPLIGAEIMGLICHCQATPHHQOESIDTFOQLAOLAIQ
    OASIVELOQGLAERHQELGTOSNQGLASNGELARATLKBEPFLNMSHELRPL
    NALIDMTLEGQEVFGMTNNKQFPALETTERSSHLDLINDIDVAKIEGQILQDC
    TTISVANNCESLVFIKQALQKRIQVEIKIPNLDPDLPVBERIRIQLINLVNNAVK
    FTFEAGRISLEVTQSLSDPIDSTVQNLQIAVDTGQIENINKLQFPFIQIDSANR
    OYEGGGLGALVKKIVELGGRVGLTSLGSGTTELPICIPSTPASPDRILDLTT
    PPELDSPTADEAVSIAPVILLIADNNXNITPSSYSAKQYDILLATNGOATELAT
    IHQDLITMDIQMGITDGLERIQIRLDLPQLVNPIVALTALMTGDRCLQACAND
    VLVKVKKQALATITIQDLITTKK"
    /gene="a113121"
    /complement(9684..10832)
    /note="ORF ID: a113121"
    /codon_start=1
    /transl_table=11
    /product="two-component hybrid sensor and regulator"
    /protein_id="BAB74820.1"
  
```

```

/db_xref="GI:17132216"
/translation="MNLSILVLDLDPNPFVISTLLSEQDYLHYVNSGQEAIASLN
LFQPKLILDDWMPGQDIEVCQRIKAMPQNPPIINVITALTAKEDLARCQSGADD
FISPDINALEARLHSMRKQQYDKIQTLSIKVLESTLDELGNDAATLP
HELNPGLNGVSGISILAMTDMTARIEQLLGNVDSARNLERLIKQLTYLELEL
STNQOONTASSTHPSIAAIEAILQSHAPSNNRRDLMFAIEBAEVSISARYLTIIH
ELVDNALFSRGTTIKTSQIVAGMLNVIHDFGRGWTAEQISKIGAFMFERKTYE
COQIGMGJAKLAKQTVESVGGFLVSSVYQQTTHIALPIAYSRHSS"
11116..11856
/gene="ccda"
11116..11856
/gene="ccda"
/feature="ORF ID:alr3122"
putative c-type cytochrome biogenesis protein CcdA"
/codon_start=1
/transl_table=11
/protein_id="BAB74821.1"
/db_xref="GI:17132217"
/translation="MLETLOQIYQLEKFAVNTLYSYQLTHLSLSIGLIFAGLLTSL
TPCMUSMPTITIGYGGYEAKSRWQAQSTWFALGLATLTLAGLGVAAMVGRVYQGI
GIGLPIVSIITAILMGLNLEALPQLSPFGDTNWSQDLPFGVRSYGIGLTYGLVAS
PCSTPVLASLQWANTQDLILGAVLLISYTAGYVTPILLAGTFTASIKKLELRWS
GWNPVSGALLVGFQVFLSIRPLGSP"
11860..13245
/gene="ycf44"
11860..13245
/gene="ycf44"
/feature="ORF ID:alr3123"
/transl_table=11
/product="c-type cytochrome biogenesis protein Ccs1"
/protein_id="BAB74822.1"
/db_xref="GI:17132218"
/translation="MTTNSAPTSPPMSLPKFLRFLPVLTDRLAIALLLIALL
PSISTEIOGSPAFYQSNYPHEPALFGLTWKVIQVGLDHYVTRWFLSLVLFQ
TSLACTTROLPAKTRQWKYEEPOFOKLALSAELDAGSVNSLSQILQNRKYI
FOEKDDILYARKGPIGRPIVHIGIYVILLGWSGAMTGFIAQENVPSETQVKN
IDAGLAPQGVQDMSVRNRFWIDYTPKGDIDQFYSDMVSVDNQGVQEHKIFVN
QPLRYHVTFTYDNGISGRVRLNKSPIFOLPALNLTNGGRWGTWIPKPDLSI
GVSLAKLOQMLLYIDAGKLVDTVRAGMSTQVNGVTLKVLVDVVGSTGLQIKADPGI
PIVYTFGLMLGVMSVFSHSQIWLQKGRDLYVGGKTRNAQVAFEQVEQDILLERLN
SOSATVINOQS"
complement(13326..13661)
/gene="alr3124"
complement(13326..13661)
/gene="alr3124"
/note="ORF ID:alr3124
similar to Xis1 protein"
/codon_start=1
/transl_table=11
/protein_id="BAB74823.1"
/db_xref="GI:17132219"
/translation="MDKLVKYQELVKLTLLTNVSDVSDVDQVEVOLIILDRNHYQWM
NVGQGLNRIYRCVIFHFDIKQKILWQNLNLTDRNPABELVMGVPRDILVGLQAPYK
ROYTDYGVN"
14186..14899
/gene="alr3125"
14186..14899
/gene="alr3125"

```

## Alignment Scores:

Pred. No.:	0.0553	Length:	348250
Score:	187.50	Matches:	91
Percent Similarity:	41.76%	Conservative:	61
Best Local Similarity:	25.00%	Mismatches:	141
Query Match:	9.16%	Indels:	71
DB:	1	Gaps:	21

US-09-889-746-2 (1-400) x AP003592 (1-348250)

Qy 6 TyrPheGlnTrpLeuSerLeuProLeuLeuSerValAlaValThrGlnGlnLeu----- 23

Db 223881 TATTTGAGTGAATTCAGTTACCCACCACCACTGCTGAATATTATTAATCAATCGACCA 223822

```

Qy 24 -----TyrAlaGlnProAsnGluSerLeuProThrValGlu----- 35
Db 223821 GGGGACTTAAATCCAGAGGGGTGAACCTAGCGAACAG---CCGGAATTTAGGAACCTTCT 223765
Qy 36 LeuGluProValValIleThrIleAspLysSerGlyMetAlaLeuAlaAsnArgIleThr 55
Db 223764 AGTGACGATGACAGATATTACCTAGAA-----GCCATCGCGGAACCAAAACC 223717
Qy 56 GlnMetProHisThr---ThrLysValIleTyrGluGluGlnIleGlnGluGlnAlaThr 74
Db 223716 CTGCCAGCATCTACTCCACCTTACGTTCATTCACCAAGAAGAAATTCAAAACAA----- 223663
Qy 75 GlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSerSer 94
Db 223662 GGTGCTACCACTGCTGCTGATGCTCTTAAAGAAATG---CCTGCTTTTGCCTAATGAT 223606
Qy 95 -----GlyThrThrSerAsnPheGlyGlnThrMetHisGlyArgGlnValGln--- 110
Db 223605 GTTGGTCATGGTGCAGATATACACAGGTACTTATTATTCGGGAGGCTCAATTAATCAG 223546
Qy 111 -----PheLeuLeuAsnGlyValProLeuThrGlySerArgAspIle----- 124
Db 223545 TCTGTATTTCCTATAAATGGCAGACGAATT-----AACATGATGTCAACACATATCAT 223492
Qy 125 ---SerArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSer 143
Db 223491 GGTGCAACTGACCTGAATAGCATTCCTGTAGAACTTATAGAGCGTGTGCAATTTATCTAGC 223432
Qy 144 GlyAlaThrSer---IleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThr 162
Db 223431 GGTGTGACCTCTGCTTTATATGGTTCCTCCGCTTTTGGAGAGGTGGTAAATATCATCACC 223372
Qy 163 LysSerAspLeuGluGluGluGlnPheGluThrArgIleGlyValHisGlySerLysLeu 182
Db 223371 AAAAAAGGTTATCCACACAACTAAATGACAAAGTAGTTAGAAATTTGGCTCACTCAATCTC 223312
Qy 183 SerSerGluGlyIleGlyTyrGlnValGlyGlnSerValAlaGlyValSerGluAsnGly 202
Db 223311 AATATCAACAGTTAGTTAT-----AGTGGGCGAGTGGTGGCGGCACTTATAAT 223261
Qy 203 AsnValLeuAlaArgLeuAspValAsp-----TyrArgThrThrGlyGlyValAlaPheAsp 220
Db 223260 TTTAGTTTGGAGCGGTACTTTGTTGATTAACCGTTATCGTGTCCCTGTAGGTGCA----- 223207
Qy 221 AlaAsnGlyLysArgIleAlaProGluProAlaGlnThrAspLysGln----- 236
Db 223206 -----GCAAAATCGTGACTCCCAAGGGTTTTATCA 223177
Qy 237 -----AspSerLysSerLeuSerValAsnThrAsnValAspTrpGlnLeuAspAspLys 254
Db 223176 AATGCAGATATACCTACTAGCACCTTATTGGCAATATTGGTATTAGATTAGATCAAGA 223117
Qy 255 GlnAsnIleAsnLeuAlaLeuThrHisTyrAsnAspLysGlnAspThrAspTyr----- 272
Db 223116 AACTCTTTAAGTTTAGATATCATAAATCAGCAGTCGTAGAGGCTTAGTTTATTTCGGA 223057
Qy 273 AlaProAspTyrGlyAsnArgLeuAlaValLeuPheGlyGluLysProSerLeuAsnAla 292
Db 223056 TTTCCCTTACAAAGAGATAGATT-----GACCATGATGGTTTAAATATT 223012
Qy 293 IleLysGlyLeuSerLeuSerGluGln-----ProLysThrThrLysSer 307
Db 223011 -----GGTTTATCTTGGAAAACCTCGCTCGGTAATGGCAATAAATCTTAACCTGCAACT 222958
Qy 308 ThrPheAsnIleAsnTyrHisAspAspLeuTrpGlyAsnThrIleAsnThr---Asn 326
Db 222957 ACATTGGTTATATACCAAAATATTATTCAGACCTACGCGCCCTACAGATTTTTTCAGAGGTAGA 222898
Qy 327 AlaTyrTyrArg 330
Db 222897 GAATTTTATCGC 222886

```

Search completed: December 25, 2002, 21:41:43  
Job time : 3632 secs

---

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 25, 2002, 20:32:49 ; Search time 254 Seconds  
(without alignments)  
3546.454 Million cell updates/sec

Title: US-09-889-746-2

Perfect score: 2047  
Sequence: 1 MKSHYFQWLSLPLSVAVT.....SKAEVLGRVPLNKPKALF 400

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frames\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09889746.r/unat\_20122002\_143748\_14192/app\_query.fasta\_1.583  
-LOOPEXT=0 -UNITs=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=45  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09889746 @CGN 1 1 125 @runat\_20122002\_143748\_14192 -NCPU=6 -ICPU=3  
-NO\_XIPXY -NO\_WARP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
1: N Geneseq\_101002.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2047	100.0	1203	21	AAA59347
2	527	25.7	2442	23	AA551448
3	463	22.6	3223	11	AAQ06026
4	199.5	9.7	2495	22	AAQ03054
5	192	9.4	2127	24	AAQ09153
6	182.5	8.9	4039	21	AAAI5184
7	180.5	8.8	2124	20	AAK91719
8	180.5	8.8	2316	20	AAK91690
9	180.5	8.8	2271	20	AAK91592
10	180	8.8	885	22	AAF26439
11	178.5	8.7	2289	20	AAK91566
12	171.5	8.4	1827	23	AA582047
13	152.5	7.4	2127	22	AAF26420
14	152	7.4	2169	21	AAK95216
15	152	7.4	2586	22	AAF26411
16	149	7.3	1980	22	ABA98989
17	149	7.3	18831	20	AAV81945
18	149	7.3	37895	17	AAK34137
19	149	7.3	37895	19	AAV00503
20	149	7.3	37895	21	AAK75158
21	149	7.3	37895	22	AAK47833
22	149	7.3	37913	14	AAO51128
23	148.5	7.3	10731	22	AA546247
24	147.5	7.2	2157	23	AA582105
25	147.5	7.2	2910	21	AAK71576
26	146	7.1	2127	21	AAZ54328
27	146	7.1	56485	21	AAK81476
28	146	7.1	349980	21	AAK21612
29	146	7.1	837096	21	AAK81489
30	145	7.1	679	23	AAK59217
31	144.5	7.0	8907	21	AAZ36926
32	143.5	7.0	9811	20	AAK84323
33	143.5	7.0	9811	20	AAK84323
34	142.5	7.0	10353	22	AA546270
35	140	6.8	1914	22	ABQ90127
36	138	6.7	8266	20	AAZ31949
37	136	6.6	8041	17	AAK10105
38	135.5	6.6	2139	22	ABA88835
39	135.5	6.6	7315	22	ABA88834
40	135	6.6	2727	23	AAK93059
41	135	6.6	2808	23	AAK88838
42	134	6.5	2448	23	AAK51469
43	134	6.5	3438	18	AAK95247
44	133.5	6.5	686	22	AAK27475
45	133.5	6.5	686	23	ABK43834

## ALIGNMENTS

RESULT 1  
AAA59347  
ID AAA59347 standard, DNA, 1203 BP.  
AC AAA59347;  
XX  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE DNA encoding a BASB047 polypeptide.  
XX  
XX BASB047; BASB054; BASB068; BASB069; vaccine; bacteremia; meningitis;  
XX Neisseria meningitidis disease; upper respiratory tract infection;  
XX invasive bacterial disease; ss.  
XX  
XX Neisseria meningitidis.  
XX  
XX  
FH Key Location/Qualifiers  
CDS 1..1203

/\*tag= a  
/product= "BASB047"

WO200043519-A2.

27-JUL-2000.

19-JAN-2000; 2000WO-EP00428.

22-JAN-1999; 99GB-0001368.

28-JAN-1999; 99GB-0001944.

29-JAN-1999; 99GB-0002086.

15-FEB-1999; 99GB-0003417.

16-FEB-1999; 99GB-0003535.

(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

Ruelle J;

WPI; 2000-505839/45.

P-PSDB; AAB07761.

Neisseria meningitidis BASB047, BASB054, BASB068, and BASB069 proteins, useful for treating N. meningitidis infections, bacteremia, and meningitis

Claim 11; Page 92; 103pp; English.

The specification describes Neisseria meningitidis polypeptides designated BASB047, BASB054, BASB068, and BASB069. The polynucleotide sequences can be used to create a vector to transform a host cell. The host cell can be used to produce the polypeptide. The polynucleotides and polypeptides can be used in vaccine compositions. The polynucleotides can be used in compositions for preparation of medicaments. The antibodies can also be used in a composition for treating humans with Neisseria meningitidis disease. The diseases that can be treated include upper respiratory tract infection, and invasive bacterial diseases such as bacteremia and meningitis. The nucleic acid sequences can be used as probes in the diagnosis of Neisseria meningitidis disease. The present sequence encodes a BASB047 polypeptide.

SQ Sequence 1203 BP; 359 A; 261 C; 263 G; 320 T; 0 other;

#### Alignment Scores:

Pred. No.:	2,66e-188	Length:	1203
Score:	2047.00	Matches:	400
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	21	Gaps:	0

US-09-889-746-2 (1-400) x AAA59347 (1-1203)

Qy	1	MetArgHisSerHisTyrPheGlnTrpLeuSerLeuProLeuLeuSerValAlaValThr	20
Db	1	ATGCGTCATCTCCACATATTTTCAATGGTATCTTTGCCCTTTTACTAAGTGTGGCAGTA	60
Qy	21	GlnGlnLeuTyrAlaGlnProAsnGluSerLeuProThrValGluLeuGluProValVal	40
Db	61	CAGCAGTTGTACGCTCAACCCCAATGATGATATACCAACGGTTGAATTAGAGCTGTGGTT	120
Qy	41	IleThrIleAspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThr	60
Db	121	ATTACCATTTGATAAGCGGTATGGCACTTGCCATTCGATATCACCAATGCCCATACC	180
Qy	61	ThrLysValIleTyrGluGluGlnIleGlnGluGlnAlaThrGlySerArgGlnLeuAla	80
Db	181	ACCAAGTTATTTATGACAGACAAATTCAGAGCAGCAACAGGCTCTCGACAGCTTGCC	240
Qy	81	AspValMetAlaGlnLeuIleProSerLeuGlyValSerSerGlyThrThrSerAsnPhe	100
Db	241	GATGTGTGGCACACGCTCATTTCCAGTTTGGGGTAAAGTAGTGGCACTACCACTTA	300

Qy	101	GlyGlnThrMetHisGlyArgGlnValGlnPheLeuLeuAenGlyValProLeuThrGly	120
Db	301	GGGCAAAACCATGCGCGTCTGCAAGTGCATTTTGTAAATGGCGCTTTCACAGGT	360
Qy	121	SerArgAspIleSerArgGlnLeuAsnSerIleAsnProAenGlnValAlaAerGlu	140
Db	361	TCGGGAGACATCTCTAGACAGCTTAATAGTATCAATCCCAATCAAGTGGCTAGAA	420
Qy	141	ValLeuSerGlyAlaThrSerIleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIle	160
Db	421	CTTTTATCAGGAGCAACAGTATTTATGGGTCTGGAGCAACAGCGGTTTTCATTAATC	480
Qy	161	ValThrLysSerAspLeuGluGluGlnPheGluThrArgIleGlyValHisGlySer	180
Db	481	GTTTACTAGTCTGATTTTGGAAAGAGGAGCAATTTGAAACCCCGCATCGGTGATG	540
Qy	181	LysLeuSerSerGluGlyIleGlyTyrGlnValGlnSerValAlaGlyValSerGlu	200
Db	541	AAATTTATCCAGTGAAGGTATCGGTTTATCAGGTAGGTGAGAGTGTGAGCGCAA	600
Qy	201	AsnGlyAsnValLeuAlaAerGluLeuAspValAspTyrArgThrThrGlyGlyVal	220
Db	601	AATGGTAAATGTCCTTGCACGACTTGATGTCCACTATCCACCAAGAGGGGCA	660
Qy	221	AlaAsnGlyLysArgIleAlaProGluProAlaGlnThrAspLysGlnAspSerLys	240
Db	661	GCTAACGGTAAACGCATCGCTCTCGAGCTGCCCAAACTGATAAGCAAGACAG	720
Qy	241	LeuSerValAenThrAsnValAspTrpGlnLeuAspLysGlnAenIleAsnLeuAla	260
Db	721	CTAAGTGTCAATACAAATGTTGATGGCACTTGACCAAGCAAAATATCAATCTG	780
Qy	261	LeuThrHisTyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArg	280
Db	781	TTGAGCATATATACGCAAAACAGATACCGATTTATGCACCTGATTTATGGTA	840
Qy	281	AlaValLeuPheGlyGluLysProSerLeuAenAlaLysGlyLeuSerLeuSerGlu	300
Db	841	CGCGTGTGTTGGAGAAAAGCCTTCATTAATGCCATCAAAAGGCTTATCATTA	900
Qy	301	GlnProLysThrThrLysSerThrPheAsnIleAenTyrHisAspAspLeuTrpGly	320
Db	901	CAGCAAAACCCCAAAAGCACCTTTAATATCACTCATCATCATGATTTGTGGGG	960
Qy	321	AsnThrIleAenThrAsnAlaTyrTyrArgArgGluLysGlyArgPheTyrProPhe	340
Db	961	AACACCATCAATACCAATGCTTATTTATCGCAGAGAGAAAGGACAGATTTTAT	1020
Qy	341	AlaProPheSerIleAlaLysAlaLeuProIleLeuGlnSerMetAsnLeuProSer	360
Db	1021	GCCCCGTTTTCGATCGCCAAAGCCCTGCTATTTTACAAAGCATGAATTTGCC	1080
Qy	361	ThrLeuAspAlaTyrThrLysAlaProGlnAlaArgAlaTyrGlyValLeuGlnSer	380
Db	1081	ACTTTGGATGCTTATACCAAGCTCCCAAGCTCGCGCTATGGGTGTTTACAAT	1140
Qy	381	SerLysAlaGluValLeuGlyValArgValProAsnLeuAenLysProLysArgAla	400
Db	1141	TCTAAGGACAGAGGTACTAGGGGTGCTCCCTAATTTGAATAGCCCAAGAGCC	1200

#### RESULT 2

AAAS51448

ID AAAS51448 standard; DNA; 2442 BP.

XX

AC AAAS51448;

XX

DT 13-FEB-2002 (first entry)

XX

DE Pseudomonas aeruginosa DNA for cellular proliferation protein #33.

XX

KW Antisense; ds; prokaryotic cellular proliferation gene;

XX

KW antibiotic; antibacterial; drug design.

XX OS Pseudomonas aeruginosa.  
XX PN WO200170955-A2.  
XX PD 27-SEP-2001.  
XX PF 21-MAR-2001; 2001WO-US09180.  
XX PR 21-MAR-2000; 2000US-191078P.  
XX PR 23-MAY-2000; 2000US-206848P.  
XX PR 26-MAY-2000; 2000US-207727P.  
XX PR 23-OCT-2000; 2000US-242578P.  
XX PR 27-NOV-2000; 2000US-253625P.  
XX PR 22-DEC-2000; 2000US-257931P.  
XX PR 16-FEB-2001; 2001US-269308P.  
XX PA (ELIT-) ELITRA PHARM INC.  
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GU,  
XX PI Yamamoto RT, Xu HH,  
XX DR MPI: 2001-611495/70.  
XX DR P-PSDB; AAU33589.  
XX PT New polynucleotides for the identification and development of  
XX PT antibiotics, comprise sequences of antisense nucleic acids -  
XX PS Claim 27; Seq ID No 4030; 51pp; English.  
XX CC The invention relates to antisense inhibitors of genes essential to  
XX CC prokaryotic cellular proliferation, their use in identifying the  
XX CC genes, their use in the discovery of novel antibiotics, the essential  
XX CC genes themselves and the encoded proteins. The prokaryotes used are  
XX CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella  
XX CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The  
XX CC invention is also useful for the identification of potential new targets  
XX CC for antibiotic development. The antisense nucleic acids can also be used  
XX CC to identify proteins used in proliferation, to express these proteins,  
XX CC and to obtain antibodies capable of binding to the expressed proteins.  
XX CC The proteins can be used to screen compounds in rational drug discovery  
XX CC programmes. The antisense nucleic acid sequence is also useful to screen  
XX CC for homologous nucleic acids which are required for cell proliferation in  
XX CC a wide variety of organisms. The present sequence encodes an  
XX CC essential prokaryotic cellular proliferation protein.  
XX CC Note: The sequence data for this patent did not form part  
XX CC of the printed specification, but was obtained in electronic  
XX CC format directly from WIPO at  
XX CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX SQ Sequence 2442 BP; 459 A; 916 C; 717 G; 350 T; 0 other;  
XX  
XX Alignment Scores:  
XX Pred. No.: 5,83e-41 Length: 2442  
XX Score: 527.00 Matches: 137  
XX Percent Similarity: 51.12% Conservative: 68  
XX Best Local Similarity: 34.16% Mismatches: 144  
XX Query Match: 25,74% Indels: 152  
XX DB: 23 Gaps: 9  
XX  
XX US-09-889-746-2 (1-400) x AAS51448 (1-2442)  
XX  
XX QY 2 ArgHisSerHisTyrPheGlnTrpLeuSerLeuProLeuSerValAlaVal---Thr 20  
XX DB 280 AGGAGCAACATACCTTC-----ATGTCGTTCCGCGCAGGAGCCGCGCGCT 333  
XX QY 21 GlnGlnLeuTyrAlaGlnProAsnGlnSerLeuProThrValGlnLeuGlnProValVal 40  
XX DB 334 ACCGACAGCCGCGAGCAACG-----CTGGACATGAGGACGATGAA 375  
XX QY 41 IleThrIleAspIleSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThr 60  
XX DB 376 ATACCGGCTTCGGCAGCAGACGAGCTGTCTCCGCAAGCGCCAG-----423

QY 61 ThrIleValIleTyrGlnGlnGlnIleGlnGlnAlaThrGlySerArgIleuAla 80  
DB 424 TCACCGCTATGAGCATGCCCACTGAGAGGCTGCGCAGGGCTCCGACGCTGGC 483  
QY 81 AspValMetAlaGlnLeuIleProSerLeuGlyValSerSerGlyThrThrSerAsn 100  
DB 484 ACCGTCTGCGCCAGAGCGGCTGGGATGTCGATTCGACGCGACATCCGAAATAC 543  
QY 101 GlyIleThrMetHisGlyArgGlnValGlnPheLeuLeuAsnGlyValProLeuThrGly 120  
DB 544 GGCACAGCCCTGCGCGGCGCAGCATGCTGTGATGTGACGGCTGCGGAAACCC 603  
QY 121 SerArgAspIleSerArgIleuLeuAsnSerIleAsnProAsnGlnValAlaArgIleGly 140  
DB 604 AACCGGACCTCTCGCGCAACTGCGCAACATGACCCGCCCTGATCGAGCGATCGAG 663  
QY 141 ValLeuSerGlyAlaThrSerIleTyrGlySerGlyValThrGlyLeuIleAsnIle 160  
DB 664 GTCATCCGTGGCAGCAGCGCCATCTACGCGAGCGCGCACCGCGGATCATCTCCATC 723  
QY 161 ValThrIleSerAspLeuGlnGlnGlnIleGlnIlePheGlnThrArgIleGlyValHisGly 179  
DB 724 ACCACCGCTCGCGCGCGCGAGAACCGCGGAAACCGGCTCGACGCCCATCTGCCG 783  
QY 180 ---SerIleLeuSerSerGlnGlnIleGlyTyrGlnValGlyIleSerValAlaGlyVal 198  
DB 784 CTACCCCGCTGCGCGCAGCATGCTCGCGCGCGCATCTTCAGCAATATCTCGCGGC 840  
QY 199 SerGlnAsnGlyAsnValIleuAlaArgLeuAspValAspTyr-----ArgThr 214  
DB 841 -----TCCCTGGGGCGCTCGACTATCTTCCTGCACTTCGCGACCCGCCAC 885  
QY 215 ThrGlyValAlaPheAspIleAsnGlyValArgIleAlaProGlnLeuProAlaGlnThrAsp 234  
DB 886 GTCGGGCTTCTTACACGCCCATGCGACCGCATGCCCGGAAACCGACCGCGGCGC 945  
QY 235 LysGlnAspSerIleSerLeuSerValIleThrAsnValAspTrpIleuAspAspIle 254  
DB 946 CTCTTGACTCGAAGCTCTACAAATCGCGCGGCAAGCTCGCTCGCATCGACGAGAAC 1005  
QY 255 GlnAsnIleAsnLeuAlaLeuThrHisTyrAsnAspIleGlnAspThrAspTyrAlaPro 274  
DB 1006 CAGCGGTGACGCTCGCGCTCAGCGCATGACGCGCGCAGGACCGCATCTCGCCACC 1065  
QY 275 AspTyrGlyAsnArgLeuAlaValLeuPheGlyGlyIleProSerLeuAsnAlaIleIle 294  
DB 1066 GAC-----CCGAGGGGTGCGCAGGCTGCGCGGCTCGCTCGCGCAACCGATCAAA 1119  
QY 295 GlnLeuSerLeuSerGlnGlnProIleThrThrIleSerThrPheAsnIleAsnTyrHis 314  
DB 1120 GGCCTGAGCTGAGCAGCAGCAACCGCATCCGACACCTCGCGCAACCTCGAGTACGAG 1179  
QY 315 HisAspAspLeuTrpIleAsnThrIleAsnThrAsnAlaTyrTyrArgArgIleuGly 334  
DB 1180 AACCTGACATCTCGCGAGCGCGCTCGCGGACGCTTACATACCGCATATTTCACC 1239  
QY 335 ArgPheTyrProPheValAlaProPheSerIleAlaValAlaLeuProIleLeuGlnSer 354  
DB 1240 CGCTTCATCTCGTTCCAGCC-----1260  
QY 355 MetAsnLeuProSerAlaThrLeuAspAlaTyrThrIleAlaProGlnAlaArgAlaTyr 374  
DB 1261 -----CGGCGCTCTCCACCCCGCGCGCG 1284  
QY 375 GlyValIleGlnSerGlnSerIleValGlnValIleGlyArgValProAsnLeuAsnIle 394  
DB 1285 AATGTGACCAATCATGACGAAGACGAAAGTGTTCGAGCGCGCTGACCGCTCGGCACG 1344  
QY 395 Pro 395  
DB 1345 CCG 1347

## RESULT 3

AAQ06026  
 ID AAQ06026 standard; DNA; 3223 BP.  
 AC AAQ06026;  
 DT 21-JAN-1991 (first entry)  
 XX Iut A iron-regulated outer membrane protein from plasmid 15972.  
 XX  
 XX Vaccines; septicaemia; iron-regulated outer-membrane protein; IROMPS;  
 XX Iut A; ss.

Key Location/Qualifiers  
 CDS 365..2564  
 FT /tag= a  
 FT /product=lut A  
 FT misc\_feature 284..286  
 FT /tag= b  
 FT /label=end of Iut D  
 FT 2891..0  
 FT /tag= c  
 FT /label=start of IS 1

EP389347-A.

26-SEP-1990.

19-MAR-1990; 90EP-0400734.

20-MAR-1989; 89FR-0003626.

(RHON-) RHONE MERIEUX.

PI Audonnet JC, Bruneau P;

DR WPI; 1990-292251/39.

Vaccines against septicaemic bacteria - contg. iron-regulated  
 outer-membrane proteins

PS Disclosure; Page 26-29; 46pp; French.

XX The gene Iut A is entirely located in the Bat E II-Bat E II  
 CC fragment (3.2 kb) of AAP15972. Comparison with pColV-K30  
 CC (99.77% homology) shows the sequence is highly conserved.  
 CC 4 nucleotide inserts alter the reading frame. The primary amino acid  
 CC structure is hereby enlarged by 8 amino acids.  
 CC A vector for expression is obtained by isolating a fragment contg. the  
 CC Iut A gene from bacterial plasmid or cloning the fragment, and inserting  
 CC it into an expression vector in which the ORI is under control of the  
 CC tac promoter and Iut A is under control of the strong "P<sub>l</sub>" promoter  
 CC with a temp. sensitive repressor, e.g. pGT1001. Bacteria  
 CC transformed are cultured at less than 32 degrees Celcius, then raised to  
 CC 42 degrees for induction of expression.

XX Sequence 3223 BP; 803 A; 859 C; 828 G; 733 T; 0 other;

## Alignment Scores:

Pred. No.:	1.3e-34	Length:	3223
Score:	463.00	Matches:	134
Percent Similarity:	45.14%	Conservative:	61
Best Local Similarity:	31.02%	Mismatches:	177
Query Match:	22.62%	Indels:	60
DB:	11	Gaps:	9

US-09-889-746-2 (1-400) x AAQ06026 (1-3223)

OY 6 TyrPheGlnTrp-----LeuSerLeuProLeuSerValAlaValThrGln 21  
 DB 383 TATACGCTTTGGGCTCTCAACCACTGCTTCTTACCATGATGGCGCCAGCAGTGGCTCAA 442  
 OY 22 GlnLeuTyrAlaGlnProAnGluSerLeuProThrValGluLeuGluProValValIle 41

Db	443	CAA	-----	445
OY	42	ThrIleAspLysSerGlyMetAlaLeuAlaAsnArg	-----	IleThrGlnMet 57
DB	446	ACCGATGATGAACGTTTCGTGTCGTGTCGCAACCGCAGCAATCGCAGTAGCGGAGATG 505		
OY	58	ProHisThrThrLysValIleTyrGluGluGlnIleGlnGluGlnAlaThrGlySerArg 77		
DB	506	CGCGAAACCACTGGTGTATCGAAACCGCCGAACTGGAACAGCAGATTCAGGGCGGCAAA 565		
OY	78	GlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSerSerGlyThrThr 97		
DB	566	GAGCTTAAAGACCGCACTGGCTCAGCTGATCCCTGGCTTGACGTGACGACGGCGGCGGC 625		
OY	98	SerAsnPheGlyGlnThrMetHisGlyArgGlnValGlnPheLeuLeuAsnGlyValPro 117		
DB	626	ACCAACTACGGTATGAATGTGCGTGGCGCGCGCTGGTGGTGGTGGTGGTGGTGGTGG 685		
OY	118	LeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAsnProAsnGlnValAla 137		
DB	686	CTCAACTCTTCACGTACCGACGCGCACTGGACTCTATAGATCCTTTTAAATATGCAC 745		
OY	138	ArgIleGluValLeuSerGlyValThrSerIleTyrGlySerGlyAlaThrGlyGlyLeu 157		
DB	746	CATATTGAAGTGATCTCCGGTGGCAGCTCCTGTACGGCGCGCGCAGTACCGGTGGCGCTG 805		
OY	158	IleAsnIleValThrLysSerAspLeuGluGluGlnPheGluThrArgIleGlyVal 177		
DB	806	ATCAACATCGTGACCAAAAAGGCGCGCAACCATGATGGAGTTTGAGCGTGGCACC 865		
OY	178	HisGlySerLysLeuSerSerGluGlyIleGlyTyrGlnValGlyGlnSerValAlaGly 197		
DB	866	AAAAGTGGCTTTAGCAGCAGTAAAGATCACCAGTAAACGCAATTGCCGAGCTGTCTCCGCG 925		
OY	198	ValSerGluAsnGlyAsnValLeuAlaArgLeuAspValAspTyrArgThrGlyGly 217		
DB	926	GGAATGAG-----CATATCTCCGGAGCTCTTCCGGGCGCATATCAGAAATTTGGCGCGC 979		
OY	218	AlaPheAspAlaAsnGlyLysArgIleAlaProGluProAlaGlnThrAspLysGlnAsp 237		
DB	980	TGGTTTGACGGTAAACGGCGATGCCACCTTGTGATAACACCCAGCGCGCTGCAGTAC 1039		
OY	238	SerLysSerLeuSerValAsnThrAsnValAspTrpGlnLeuAspAspLysGlnAsnIle 257		
DB	1040	TCCGATCCGGCTGGACATCATCGGAACCTGGTACGCTGAACATCGATATCCCGCAGCTT 1099		
OY	258	AsnLeuAlaLeuThrHisTyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGly 277		
DB	1100	CAGTTGATCACACAGTACTATAAAGCCAGCGCGCGCAGCAGTACGGCTTAATCTCGGG 1159		
OY	278	AsnArgLeuAlaValLeuPheGlyGlyLysProSerLeuAsnAlaIleLysGlyLeuSer 297		
DB	1160	AAAGCTTCTCTGCCATCAGAGGGGACGACGCGCATTC---GTCACTAACGGCTGAAT 1216		
OY	298	LeuSerGluGlnProLysThrThrLysSerThrPheAsnIleAsnTyrHisHisAspAsp 317		
DB	1217	TCCGACCGCTATTCCCGGCACCTGAGCGGCAATTTGATCAGCCTGCAGTACTCTGCACAGCGCT 1276		
OY	318	LeuTrpGlyAsnThrIleAsnThrAsnAlaTyrTyrArgArgGluLysGlyArgPheTyr 337		
DB	1277	TTTCTGGGACAGGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1336		
OY	338	ProPhe-----ValAlaProPheSerIleAlaLysAlaLeu 349		
DB	1337	CGGTTCCCGACGGTAAATGCGAACAAACAGGTGACGGCTTCTTCTCTCTGCACAGCAGGAC 1396		
OY	350	ProIleLeuGlnSerMetAsnLeuProSerAlaThr----- 361		
DB	1397	ACCGACCACTAGTGGGATGAAACTGACTCTGAACAGCAACCGATGGACGGCTGGCAATC 1456		
OY	362	-----LeuAspAla-----TyrThrLysAla 368		

Db	1457	ACCTGGGGGGCTGTGATGCTGATCATGACCGCTTACTCCACACAGATGTTCTTCGACCTG	1516
Qy	369	ProGlnAlaArgAlaTyrGlyValLeuGlnSerGluSerIleValIleuGlyArg	388
Db	1517	GCTCAGGAGACGCTTCTCCGAGGGCTGACACAGAG---AAGATTTCACCAACCGGGCGC	1573
Qy	389	ValProAenLeuAenIlyProIlyArgAlaLeuPhe	400
Db	1574	TATCCGTGTGTATGACATCACCAACCTGGCGCTTC	1609
RESULT 4			
AD	AAD03054	standard; DNA; 2495 BP.	
AC	AAD03054;		
XX			
DT	13-JUN-2001	(first entry)	
DE			
XX		ironEc gene fragment from an extraintestinal isolate of Escherichia coli	
KW	ironEc: extracytoplasmic protein; immunogen; vaccine; EI; UTI;		
KW	immunotherapy; extraintestinal infection; urinary tract infection;		
KW	meningitis; pneumonia; intra-abdominal infection; antibiotic; de.		
XX			
OS	Escherichia coli.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	321..2495	
FT		/*tag= a	
FT		/product= "B. coli ironEc extracytoplasmic protein"	
FT		/transl_except= (pos:678..686, aa:Val-Tyr)	
FT		/note= "insertion of 3 bases alters the reading frame"	
FT		/transl_except= (pos:723..728, aa:Asn-Arg-Trp)	
FT		/note= "There is an apparent deletion of one codon	
FT		corresponding to Arg, which alters the reading frame;	
FT		the coding region does not include stop codon"	
FT		/partial	
FT	sig_peptide	321..392	
FT		/*tag= b	
FT	mat_peptide	393..2495	
FT		/*tag= c	
FT		/product= "Mature B. coli ironEc extracytoplasmic	
FT		protein"	
XX			
XX	WO200121636-A1.		
XX			
PD	29-MAR-2001.		
XX			
PE	22-SEP-2000; 2000MO-US26117.		
XX			
PR	22-SEP-1999; 99US-0155621.		
XX			
PA	(UYNY ) UNIV NEW YORK STATE RES FOUND.		
XX			
PI	Russo T, Carlino U;		
XX			
DR	WPI; 2001-244936/25.		
DR	P-PSDB; AAY72914.		
XX			
PT	Novel isolated ironEc polynucleotide from extraintestinal isolate of		
PT	Escherichia coli useful as vaccine for treating or preventing		
PT	extraintestinal infections caused by extraintestinal pathogenic		
PT	Escherichia coli		
XX			
PS	Claim 5; Page 37-38; 44pp; English.		
XX			
CC			
CC	The invention relates to the identification of ironEc gene,		
CC	from an extraintestinal isolate of Escherichia coli. This		
CC	gene is expressed in increased amounts in human urine and is		
CC	identified by transposon (Tnpho) mutagenesis. ironEc gene		
CC	encodes an extracytoplasmic protein. This gene can be used as		
CC	an immunogen in vaccine formulations. The recombinant vector		
CC	comprising nucleotide sequence encoding one or more antigenic		

CC	epitopes of irokhec is useful for diagnostic and immunotherapeutic			
CC	purposes. The irokhec antigenic peptide is useful for treating or			
CC	preventing extraintestinal infections (EIs) caused by extra-			
CC	intestinal pathogenic E. coli (ExPEC). The EIs include urinary			
CC	tract infection (UTI), meningitis, intra-abdominal infection			
CC	and pneumonia.			
CC	The present sequence is irokhec gene, from an extraintestinal			
CC	isolate of Escherichia coli.			
XX				
50	Sequence 2495 BP, 686 A; 557 C; 632 G; 620 T; 0 other;			
	Alignment Scores:			
	Pred. No.:	2,72e-09	Length:	2495
	Score:	199.50	Matches:	84
	Percent Similarity:	41.94%	Conservative:	59
	Best Local Similarity:	24.63%	Mismatches:	135
	Query Match:	9.75%	Indels:	63
DB:		22	Gaps:	16
US-09-889-746-2 (1-400) x AAD03054 (1-2495)				
Qy	9	TrypLeuSerLeuProLeuLeuSerValAlaValThrGlnGlnLeu-----	Tyr	24
Db	342	TGG---TCGGTAACTGGCTCTCTTATGGGTGAATAGCCAGGTATGACCAAAATAC		398
Qy	25	AlaGlnProkangIuSerLeuProThrValGlnLeuGluProValIleThrIleAsp		44
Db	399	TCCGACGATGATTAATGAC-----GACACTCTGTGTGTGGA-----		434
Qy	45	LySerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProIleThrThyValIle		64
Db	435	-----GCCACCGCTGAGCAGGATTAATAACAGACCGGGGGTGTGGTATT		482
Qy	65	TyrGlnGlnIleGlnGlnGlnAlaThrGlySerArgGlnLeuAlaAspValMetAla		84
Db	483	ACCAGCAGGATATT-----AAAAGACCCCTCCGGTAAACGACTTTCAGATTAAT		539
Qy	85	GlnLeuIleProSerLeuGlyVal-----SerSerGlyThrThySerAspPhe		100
Db	540	AAATG---CCTGTGTATTCTTACCGCATAGCGCTCGGCGACACCGCGTAAATAC		596
Qy	101	GlyIleThr-----MetIleGlyArgGlnValGlnPheLeuLeuAsnGlyVal		116
Db	597	CGCCAGATCATATTGCTGTGTATGGGCGCGGAAACACTTAATTTATGATGATGTGA		656
Qy	117	ProLeuThr-----GlySerArgAspIleSerArg		126
Db	657	CCGGTACGTCACGTAACTCCGTCGTTATAGCTGGCGTGGGAGACCGTAAACCCGCG		716
Qy	127	GlnLeuAsnSerIleAsnProkangIleValAlaArgIleGlnValLeuSerGly---Ala		145
Db	717	GACACCAACTGGGTGCCACCGAACAAGGTGACGTTGAATGATATCCGCGCCCTCG		776
Qy	146	ThrSerIleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrIlySerAsp		165
Db	777	GGCGCGCGCTACGGTTGGGGGCGCGCGGGGGGTGTGAACATCAATACAAAGTCCC		836
Qy	166	LeuGlnGlnGlnGlnPheGlnThrArgIleGlyValHisGlySerIlyLeuSerSerIu		185
Db	837	ACCAACACCTGGACCGGTGCTGTGTTATACCAACCAACCGCGGAAGTACCAAG		896
Qy	186	GlyIleGlyTyrGlnValGlyGlnSerValAlaGlyValSerGluAsnGlyAAsnValLeu		205
Db	897	GGCGCTACGGCTCGCGGCATTTACGCTTATGTAGTGGCCCTCTGGCTGTATGCTCTTAC		956
Qy	206	AlaArgLeuAspValAspTyr---ArgThrThrGlyGlyAlaPheAspAlaAsnGlyLys		224
Db	957	ACGGGTGTGTATGTGTAACCTGAATAAAGCGATGCTGACAGTTGGAGATTAAT-----		1010
Qy	225	ArgIleAlaProGluProAlaGlnThrAspLys-----GlnAspSerIlySer		240
Db	1011	-----TCTCCGGTGGTACGAAAACGACGCCGGGCGATGAAGGGGTATGTAAC		1058

Qy 241 LeuSerValAsnThrAsnValAspTrpGlnLeuAspAspLysGlnAsnLeuAla 260  
 Db 1059 AAAGATATTAACGGGTTGCTCGTGGAAATTAATCCGACGAGATTCTCGATTTCGAA 1118  
 Qy 261 LeuThrHisTyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeu 280  
 Db 1119 GTCGATAT-AGCCGCCAGGGGATATCTATGCGGGGATACCGACGACAGTTCT 1172  
 Qy 281 AlavalLeuPheGlyGluLysProSerLeuAsnAlaLeuLysGlyLeuSerLeuSerGlu 300  
 Db 1173 TCCAGTGCAGTTACCGAA-AGCCTGGCAAAA 1202  
 Qy 301 GlnProLysThrThylsSerThrPheAsnLeuAsnTyr-HisHisAspAspLeu 318  
 Db 1203 TCCGGCAAAAGACGACGACCGCTGTACCGACAGAAATTATGGCATTACGCATAATGGTATC 1262  
 Qy 319 TTP 319  
 Db 1263 TGG 1265  
 RESULT 5  
 ID ABQ90153 standard; DNA; 2127 BP.  
 XX AC ABQ90153;  
 XX DT  
 XX DE  
 XX DE M. capsulatus gene #138 for DNA array.  
 XX KW Micro array; gene; ds; differential expression; gene expression.  
 XX OS  
 XX OS Methylococcus capsulatus.  
 XX PN WO200255655-A2.  
 XX PD 18-JUL-2002.  
 XX PX 14-JAN-2002; 2002WO-NO00019.  
 XX PR 12-JAN-2001; 2001NO-0000235.  
 XX PR 12-JAN-2001; 2001NO-0000239.  
 XX (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.  
 PA (TIGR-) TIGR.  
 XX PI Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;  
 PI Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS;  
 PI Salzberg SL;  
 XX WPI; 2002-557818/59.  
 PT Novel DNA array useful for determining differential expression of  
 PT Methylococcus capsulatus genes, comprises polynucleotides or  
 PT oligonucleotides representative for a selective number of Methylococcus  
 PT capsulatus genes -  
 XX Claim 19; Page 118-119; 678pp; English.  
 CC The invention relates to a novel DNA array giving a representation of a  
 CC number of Methylococcus capsulatus genes. The method of the invention is  
 CC useful for determination of the differential expression of the genes of  
 CC M. capsulatus, and for studying gene expression on a genomic scale and in  
 CC gene expression assays of M. capsulatus genes. The sequences shown in  
 CC ABQ90016-ABQ91855 represent M. capsulatus genes for use in arrays of the  
 CC invention.  
 XX SQ Sequence 2127 BP; 427 A; 670 C; 663 G; 367 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 1,168-08 Length: 2127  
 Score: 192.00 Matches: 99  
 Percent Similarity: 38.01% Conservative: 69

Best Local Similarity: 22.40% Mismatches: 172  
 Query Match: 9.38% Indels: 102  
 DB: 24 Gaps: 16  
 US-09-889-746-2 (1-400) x ABQ90153 (1-2127)  
 Qy 4 SerHisTyrPheGlnTrpLeuSerLeuProLeuLeuSerValAlaValThrGln----- 21  
 Db 22 TCTTCGTGCTTTCATCGTGTGGCGTCCGCTCATCCGCCGAGGTTTCGAGCCCGAG 81  
 Qy 22 -----GlnLeuTyrAlaGlnProAsnGluSerLeuProThr 33  
 Db 82 GGCAGCAGGTCCTCGGGGGAAGGATCGGCCCTCCGTCGGGAGCCGGCTCCC--- 138  
 Qy 34 ValGluLeuGluProValValIleThrIleAspLysSerGlyMetAlaLeuAlaAsnArg 53  
 Db 139 GTGAGCTCGAACCCTGCTCATCAGC-----AGTCCTTGGAGAAAG 183  
 Qy 54 IleThrGlnMetProHisThrThrLysValIleTyrGluGluGlnIleGlnGluGlnAla 73  
 Db 184 GTCTCCGAAATGGCGCGCGCGGTGACCGTGTGACGGGACGCAATTCGCCCGGAAATC 243  
 Qy 74 ThrGlySer-----ArgGlnLeuAlaAspValMetAlaGlnLeuIlePro 88  
 Db 244 GGGGCGCAGTCGTGTAAACCCCTCAAGCAGGAGCCGCGTGACCGGAA----- 294  
 Qy 89 SerLeuGlyValSerSerGlyThrThrSerAsnPheGlyGlnThrMetHisGlyArgGln 108  
 Db 295 TCTTTCGTCCCGGTAGGCATACCGGTGTATCCGGGGCAGACG-----GGCGCGGT 348  
 Qy 109 ValGlnPheLeuLeuAsnGlyValProLeuThrGlySerArgAspIleSerArg----- 126  
 Db 349 GTCCAGGTAATGACCAACAGCATA-----GGCACCAGCGACGCTCCAGATCACC 399  
 Qy 127 -----GlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGly 144  
 Db 400 CGGACCATGCAATGCGGTGGAGCCGCTGCGCGCCGACCGGTGGAAGTACTTGGCGGG 459  
 Qy 145 ---AlaThrSerIleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLys 163  
 Db 460 CGGCGACCTGCTGTACGCGAGCGGGCCATCGCGGCATCGTCAACGCTCTCGACAA 519  
 Qy 164 -----SerAspLeuGluGluGlnPheGlu 172  
 Db 520 CGCATTCCTCGTGTGCGGCAAGCCGATGACGCGTACCGGGGAACACGCGTTCAAC 579  
 Qy 173 ThrArgIleGlyValHisGlySerLysLeuSerSerGlyGlyIleGlyTyrGlnValGly 192  
 Db 580 TCGGTCTCTGCCAAACACCAACCAATCTCAAGGAGAGGGCGGCGCAGGATCTGGTGTC 639  
 Qy 193 GlnSerValAlaGlyValSerGluAsnGlyAsnValLeu-----AlaArg 207  
 Db 640 TATCATCTGGACGGGTTTTCGCGGATGTTAAACATCTGCACATCGCGGCGCGCGATC 699  
 Qy 208 LeuAspValAspTyrArgThrThrGlyAlaPheAspAla-----AsnGly 223  
 Db 700 GCCGAGACTCAGGCGCGCAGACCCGACCTGCTTTGCAAGCGACGCCCGATCCAGAA 759  
 Qy 224 LysArgIleAlaProGluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerVal 243  
 Db 760 TACGGCGTCATCCCAACACCGAAATCGGCCAAGGCGGTTTCGCGGGGTTTTCCTG 819  
 Qy 244 -----AsnThrAsnValAspTrpGlnLeuAspAsp----- 253  
 Db 820 GTGGCGAGCCGGGCTTCGCGGGGTGTCGATCAACTATCTCGGCAACAACATACGCA 879  
 Qy 254 -----LysGlnAsnIleAsnLeuAlaLeuThrHisTyr 264  
 Db 880 CCACCCGACGATCTCCCGGGGAGGCAATACCCGCATCAACCTGAAGGAGTCCGCGTAC 939  
 Qy 265 AsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeuPhe 284  
 Db 940 GACTTCAAGAGCGAGCTGAATGCGCGGTGATTTCTGTCGAGGCGCGTGGCATCGCTTC 999

QY 285 GlyGlyProSerLeuAlaIleGlyLeuSerLeuSerGluInProLysThr 304  
 DB 1000 GGC----- 1002  
 QY 305 ThrLysSerThrPheAsnIleAsnIleAsnIleAsnIleAsnIleAsn 324  
 DB 1003 -----CACGTCATTACCCACGACCGAGCTGATGACAGCCGGCGCC 1047  
 QY 325 ThrAsnAlaTyrTyrArgArgGlyLysGlyArgPheTyrProPheValAlaProPheSer 344  
 DB 1048 GCTCGGTCAACAAGACGACCAAGAGCCGTCTCGAAGTGTCCACCGCCGATCGGC 1107  
 QY 345 IleAlaLysAlaLeuProIleLeuGlnSerMetAsn-----LeuProSer 359  
 DB 1108 ATCTTGAGGGCGCTGTTGGGGTGCAGATGAGCTCGCGGATTTTCGAGCGCATCGAGATC 1167  
 QY 360 AlaThrLeuAspAlaTyrThrLysAlaProGlnAlaAlaArgAlaTyrGlyValLeuGlnSer 379  
 DB 1168 GGTTCGCCGACGCGCTGTGTGCCCAAGTCCCACTCGGACGTACGCGTTCGCCGTC 1227  
 QY 380 GluSer 381  
 DB 1228 GAATCC 1233  
 RESULT 6  
 AAA15184/C  
 ID AAA15184 standard; DNA; 4039 BP.  
 AC AAA15184;  
 DT 04-SEP-2000 (first entry)  
 DE DNA encoding Escherichia coli virulence proteins.  
 KM Virulence protein; tatA, tatB, tatC, tatE, mdgG, creC, recG, yggN,  
 KW eckA, iroD, iroC, iroB, mtd2, msl, vaccine, infection;  
 OS Gram negative bacterium; ss.  
 XX Escherichia coli.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..285  
 FT /tag= a  
 FT /product= "virulence protein"  
 FT /note= "encodes AAY93335"  
 FT CDS 370..1326  
 FT /tag= b  
 FT /product= "virulence protein"  
 FT /note= "encodes AAY93336"  
 PN MO200028038-A2.  
 XX  
 PD 18-MAY-2000.  
 XX  
 PF 09-NOV-1999; 99WC-GB03721.  
 XX  
 PR 09-NOV-1998; 98GB-0024569.  
 PR 09-NOV-1998; 98GB-0024570.  
 PR 17-DEC-1998; 98GB-0027814.  
 PR 17-DEC-1998; 98GB-0027815.  
 PR 17-DEC-1998; 98GB-0027816.  
 PR 17-DEC-1998; 98GB-0027818.  
 PR 13-JAN-1999; 99GB-0000708.  
 PR 13-JAN-1999; 99GB-0000710.  
 PR 13-JAN-1999; 99GB-0000711.  
 PR 28-JAN-1999; 99GB-0001915.  
 XX  
 PA (MICR-) MICROSCIENCE LTD.  
 XX  
 PI Crooke HR, Clarke BE, Everest PH, Dougan G, Holden DW, Shea JE;  
 XX Feldman RG;

DR WPI; 2000-376550/32.  
 DR P-PSDB; AAY93335, AAY93336.  
 XX  
 PT Peptide encoded by an operon including genes from Escherichia coli for  
 PT screening potential drugs, detecting virulence and treating conditions  
 PT associated with infection by a Gram negative bacterium -  
 XX  
 PS Example 10; Page 74-78; 122pp; English.  
 XX  
 CC The present sequence encodes Escherichia coli virulence proteins.  
 CC The specification describes virulence proteins which are encoded  
 CC by an operon including tatA, tatB, tatC, tatE, mdgG, creC, recG, yggN,  
 CC eckA, iroD, iroC, iroB, mtd2 or msl-16 genes obtained from Escherichia  
 CC coli K1. The virulence proteins and polynucleotides, and their vaccines  
 CC are useful for screening potential drugs, for the detection of virulence,  
 CC and for treating or preventing conditions associated with infection by  
 CC a Gram negative bacterium particularly Escherichia coli.  
 XX  
 SQ Sequence 4039 BP; 985 A; 1039 C; 942 G; 1073 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 2,3e-07 Length: 4039  
 Score: 182.50 Matches: 115  
 Percent Similarity: 37.62% Conservative: 75  
 Best Local Similarity: 22.77% Mismatches: 171  
 Query Match: 8.92% Indels: 144  
 DB: 21 Gaps: 23  
 US-09-889-746-2 (1-400) x AAA15184 (1-4039)  
 QY 9 TrpLeuSerLeuProLeuSerValAlaValThrGlnLeuTyrAla----- 25  
 DB 3528 TGG---TCGCTAAGTGTGCTCGCTGGTGGTGAATGACAGATACATGACCAATCC 3472  
 QY 26 -----GlnProAsnGluSerLeuProThrValGluLeuGluProValIleThrIle 43  
 DB 3471 TCGAGATGATATATACAGACTGTGTGTGA----- 3436  
 QY 44 AspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrLysVal 63  
 DB 3435 -----CCACCGCTGACAGGTATTAAACAGACCGCGCGTGTGTT 3391  
 QY 64 IleTyrGluGlnGlnIleGlnGlnAlaIleThrGlySerArgGlnLeuAlaAspValMet 83  
 DB 3390 ATTACAGAGAGATTT---AAAAAGACCCCTCGGTAAAGACCTTTCAGATATTATT 3334  
 QY 84 AlaGlnLeuIleProSerLeuGlyVal-----SerSerGlyThrThrSerAsn 99  
 DB 3333 CGTAAATG---CTGCGCTCAATCTTACCGCAATAGCGCTCGGCGACACCGCGTAAC 3277  
 QY 100 PheGlyGlnThr-----MetHisGlyArgGlnValGlnPheLeuLeuAsnGly 115  
 DB 3276 AACCGCCAGATGATTTCTGTGTATGGGCGGAAACACCTTAATTTTAATGATGGT 3217  
 QY 116 ValProLeuThr-----GlySerArgAspIleSer 125  
 DB 3216 GTACCGGTACGTCACTAATCCGCGCTTATAGCTGCGTGAAGCTATACCCGC 3157  
 QY 126 ArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGlnValLeuSerGly--- 144  
 DB 3156 GGTGACACCAACGCTGGGTGCACCGGAACAGGTGAGCGTATTGAAGTATCCGCGCCCT 3097  
 QY 145 AlaThrSerIleTyrGlySerGlyAlaThrGlyValLeuIleAsnIleValThrLysSer 164  
 DB 3096 GCGCGCGCGCTACGCTTGGGGGAGCGCGGGGGGTGAACATTCATCAACCAACGT 3037  
 QY 165 AspLeuGluGluGlnPheGlnThrArgIleGlyValHisGlySerLysLeuSerSer 184  
 DB 3036 CCCACCAAGACTGGCAGCGTTCGCTGTATACACCAATACGCGGGAAGTACCGAT 2977  
 QY 185 GluGlyIleGlyTyrGlnValGlyGlnSerValAlaGlyValSerGluAsnGlnVal 204  
 DB 2976 GAGGGCGTACGCGTCCGCAATTTCAGCTTAGTGGCGCTGCGCTGATGCTTT 2917

```

QY 205 LeuAlaArgLeuAspValAspTyr---ArgThrThrGlyAlaPheAspAlaGly 223
DB 2916 ACCACGCGTTTGTATGTAATCGAATAAAGCGGATGCTGACAGTTGGGATTAAT--- 2860
QY 224 LysArgIleAlaProGluProAlaGlnThrAspLys-----GlnAspSerLys 239
DB 2859 -----TCTCCGTCGTCGTAAGAAAGCGACGCGGCATGAGGGGTACGT 2815
QY 240 SerLeuSerValAsnThrAsnValAspTyrGlnLeuAspAspLysGln----- 255
DB 2814 AACAAAGATATTAAACGCGTCTCTCGTGGAATTAATCCGACGAGATCTCGATTC 2755
QY 256 -----AsnIleAsnLeuAlaLeu----- 261
DB 2754 GAAGCCGATATAGCCGCCAGGGAATCTATCGCGCGATACGCAAAACAGTTCTTCCA 2695
QY 262 -----Thr-HisTyrAsnAspLysGlnAspThrAsp-----TyrAlaPr 274
DB 2694 GTGCAGTTACCAAGACCTGGCAAAATCCGCAACAGAGACGCGCCTGTACCGACA 2635
QY 274 AspTyrGly-----AsnArgLeuAlaValle 283
DB 2634 GAATTATGCAATTACGCATAATGGTATCTGGGACTGGGACAAAGTCGTTGGTGTTA 2575
QY 283 uPheGlyGluLysProSerLeuAsnAlaIleLysGlyLeuSer----- 297
DB 2574 TTACGAGAAACCAATAATACCCGCATGAATGAAGGATTATCCGCGGTGTGCAAGGGCG 2515
QY 298 -----LeuSerGluGlnProLysThr-----LysSerThrPheAs 310
DB 2514 TATTTTAGCGGTGAAAGATTACGACCAATCGCTGAGTTCTTGGCGAACCGCGCA 2455
QY 310 nIleAsnTyrHisAspAspLeuTyrGlyAsnThrIleAsnThrAsnAlaTyrTyr 330
DB 2454 GCTTAATATCTTGAATGTATGTTGATCAAAAGCTGACCGTTGGTGACAGTGGAA 2395
QY 330 gArgGluLys-GlyArgPheTyrPro----- 338
DB 2394 CCGCATAGCTCGATGATCTTCTCTACCAGCCTGACGCGTGAATGACAGTATATCAG 2335
QY 339 -----PheValAlaProPheSerIleAlaLysAlaLeuProIleLeuGlnSer----- 354
DB 2334 CGGTATTTCTGGCTGCTGCGGATCGCAGCAGTAAATCAATCTCAATCAGTGGCT 2275
QY 355 -----MetAsnLeuProSerAlaThrL 362
DB 2274 GTATATTGAAGATAACATTGAGCGGTTCTTGGCAGCAATATCATTTCCGCGCTGGCTT 2215
QY 362 eu-----AspAlaTyrThrLysAlaProGln-----AlaArgAlaT 374
DB 2214 TGATTATCTCAGCGACTCCGCGGGGAACCTTCAGCCCACTCTGAATCTTTTCGACGAAAT 2155
QY 374 yGlyValLeu-GlnSerGluSerLysAlaGluValLeuGlyArgValProAsnLeuAsn 393
DB 2154 GGGCGATTATTCAAAGTCAAAGCAGGGGT-----TGCCCGCAACCTTTAA 2110
QY 394 LysProLys 396
DB 2109 AGCCCCAAA 2101
RESULT 7
AAX91719
ID AAX91719 standard; DNA; 2124 BP.
XX
XX AAX91719;
AC AAX91719;
XX
XX 25-AUG-1999 (first entry)
DE Porphorymonas gingivalis protein PG40 encoding DNA.
XX
XX Porphorymonas gingivalis; PG; periodontal disease; gingivitis;
KW vaccine; antigenic; ds.

```

```

XX
OS Porphorymonas gingivalis.
XX
XX W09929870-A1.
XX
XX 17-JUN-1999.
XX
XX 10-DEC-1998; 98WO-AU01023.
XX
XX 04-AUG-1998; 98AU-0005028.
XX 10-DEC-1997; 97AU-0000839.
XX 31-DEC-1997; 97AU-0001182.
XX 30-JAN-1998; 98AU-0001546.
XX 10-MAR-1998; 98AU-0002284.
XX 09-APR-1998; 98AU-0002911.
XX 23-APR-1998; 98AU-0003128.
XX 05-MAY-1998; 98AU-0003338.
XX 22-MAY-1998; 98AU-0003654.
XX 29-JUN-1998; 98AU-0004917.
XX (CSLC-) CSL LTD.
XX
XX Agius CT, Barr IG, Hocking DM, Margetta MB, Patterson MA;
XX Ross BC, Rothel LJ, Webb EA;
XX
XX WPI; 1999-385613/32.
XX P-PSDB; AAY34501.
XX
XX Antigenic Porphorymonas gingivalis peptides for preventing
XX gingivitis
XX
XX Claim 12; Page 219; 589pp; English.
XX
XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
XX Porphorymonas gingivalis (PG) polypeptide sequences given in AAY34318 to
XX CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
XX CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
XX CC activity with a vaccine mechanism of action. The PG polypeptides can be
XX CC used as vaccines especially against Porphorymonas gingivalis. Probes can
XX CC be used to detect Porphorymonas gingivalis in standard hybridisation
XX CC assays. Porphorymonas gingivalis is involved in periodontal disease
XX CC especially gingivitis.
XX
XX Sequence 2124 BP; 604 A; 567 C; 469 G; 484 T; 0 other;
SQ

```

```

Alignment Scores:
Pred. No.: 158-07 Length: 2124
Score: 180.50 Matches: 94
Percent Similarity: 40.30% Conservative: 68
Best Local Similarity: 23.38% Mismatches: 131
Query Match: 8.82% Indels: 109
DB: Gaps: 21

US-09-889-746-2 (1-400) x AAX91719 (1-2124)
QY 14 LeuLeuSerValAlaValThrGlnLeuTyrAlaGln-----ProAsnGlu 29
DB 22 CTACTATCGATTGGTATTTCAGCGCAGGCTTTTCCAAAGACGGCAACGTCGCGACAGAT 81
QY 30 SerLeuProThrValGluLeuGluProValValIleThrIleAspLysSerGlyMetAla 49
DB 82 TCGCTACGAGTACCAATCTTCACACCTCCAGGCTTATTTTACACGACGCGCGCTACT 141
QY 50 LeuAlaAsnArgIleThrGlnMetProHisThrThrLysValIleTyrGluGluGlnIle 69
DB 142 CTG-----AAAAGATACCGCCCAAGATGGAATCATCTCATCTCGCGCAACATC 189
QY 70 GlnGluGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSer 89
DB 190 AAGCAG-----TCCGGCTTTAAACAACATGACCGACATCCTC---AAGACGCAAAAGTTCG 240
QY 90 LeuGlyValSerSer-----GlyThrThrSerAnPhe-----GlyGlnThrMet 104

```





```

Db 475 TTGTGACTCGGCCCCCATCTGATGGCTCTTTCATGCGAGCTGAACGGCTTAGCAGTAAG 534
Qy 108 GlnValGlnPheLeuLeuAsnGlyValProLeuThrGlySerArgAspIleSerArgGln 127
Db 535 TATATCTCTATCTTATCGATCGTGAAGCGTGTGACGGCGATGAGGCGGTGAGGCCGAT 594
Qy 128 LeuAsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyAla---Thr 146
Db 595 TTGAGTCGTATTCTCTGATCAGATCGAAGCGATCGAATCGTGTGAAGGTCCTTCGAGT 654
Qy 147 SerIleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLysSerAspLeu 166
Db 655 TCGCTCTACGGATCCGATCCATCGCGGGGTAATCATGTGATCACAATAAAGATACG 714
Qy 167 GluGluGluGlnPheGluThr-----ArgIleGlyValHisGlySerLysLeuSerSer 184
Db 715 AATCGACTGAGTGCATATACGTACATCGCATATCGAAGTACAAAGTACGCAACCAAT 774
Qy 185 GluGlyIleGlyTyrGlnValGlnSerValAlaGlyValSerGluAsnGlyAsnVal 204
Db 775 ACTTCGCTCGATATAACATCGGTAG-----TTCAGTAGCAATACCAACTAT 822
Qy 205 LeuAlaArgLeuAspValAspTyrArgThrThrGly-----GlyAlaPheAspAla 221
Db 823 TTCCTC-----TACCATACGGATGGCTGCGAGAAATAGTCGTTCTGAATA 867
Qy 222 AsnGlyLysArgIleAlaProGluPro-----230
Db 868 AAAAGAAAAGGATCCGCGACCGGCTTCGAGGAACGTATAGAAGAACTTTTCGT 927
Qy 231 AlaGlnThrAspLysGlnAspSerLysSerLeuSer---ValAsnThrAsnValAspTrp 249
Db 928 GCACAGAAATACAGGGTGAAGCCCAATCGCTTCTTATTCGCAACTAAC-----978
Qy 250 GlnLeuAspAspLysGlnAsnIleAsnLeuAlaLeuThrHisTyrAsnAspLysGlnAsp 269
Db 979 AATCTTAGCTTCAGCCGAATGTG-----CAGTACAAATAAAGCTCAG---1020
Qy 270 ThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeuPheGlyGluLysProSer 289
Db 1021 ---ATCTTCACTCCGACTTTTCCGAAAGAGGCTATGACATGATGATTCGT---GCT 1074
Qy 290 LeuAsnAlaIleLysGlyLeu-----SerLeuSerGlu 300
Db 1075 TTGACGGCTTCACTCGGTACGAACTATCTTTTCCCAATGGTCTGCATACGCTTCTTTC 1134
Qy 301 GlnProLysThrThrLysSerThrPheAsnIleAsnTyrHisAspAspLeuTrpGly 320
Db 1135 GATGCGCTTACGATCGCTTCCGATATTTGATATCATGACAAAGACAGCAGTGTAG 1194
Qy 321 AsnThrIleAsnThrAsnAlaTyrTyrArgArgGluLysGlyArgPheTyrPro 338
Db 1195 AGCTGATCAACACCAAGGT-----CAGACCGACCAACCCACATCTCTTCG 1242

```

RESULT 9  
 AAX91592  
 ID AAX91592 standard; DNA; 2316 BP.  
 XX  
 AC AAX91592;  
 XX  
 DT 25-AUG-1999 (first entry)  
 XX  
 DE Porphyromonas gingivalis protein PG40 ORF encoding DNA.  
 XX  
 KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;  
 XX  
 OS Porphyromonas gingivalis.  
 XX  
 PN WO9929870-A1.  
 XX  
 PD 17-JUN-1999.

```

XX 10-DEC-1998; 98WO-AU01023.
XX 04-AUG-1998; 98AU-0005028.
PR 10-DEC-1997; 97AU-0000839.
PR 31-DEC-1997; 97AU-0001182.
PR 30-JAN-1998; 98AU-0001546.
PR 10-MAR-1998; 98AU-0002284.
PR 09-APR-1998; 98AU-0002911.
PR 23-APR-1998; 98AU-0003128.
PR 05-MAY-1998; 98AU-0003338.
PR 22-MAY-1998; 98AU-0003654.
PR 29-JUL-1998; 98AU-0004917.
XX (CSLC-) CSL LTD.
XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX WPI; 1999-385613/32.
DR P-PSDB; AAY34374.
XX Antigenic Porphyromonas gingivalis peptides for preventing
PT gingivitis
XX Claim 12; Page 133; 589pp; English.
XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC be used to detect Porphyromonas gingivalis in standard hybridisation
CC assays. Porphyromonas gingivalis is involved in periodontal disease
CC especially gingivitis.
XX Sequence 2316 BP; 665 A; 615 C; 505 G; 531 T; 0 other;
SQ

```

Alignment Scores:  
 Pred. No.: 1.68e-07 Length: 2316  
 Score: 180.50 Matches: 94  
 Percent Similarity: 40.30% Conservative: 68  
 Best Local Similarity: 23.38% Mismatches: 131  
 Query Match: 8.82% Indels: 109  
 DB: 20 Gaps: 21

US-09-889-746-2 (1-400) x AAX91592 (1-2316)

```

Qy 14 LeuLeuSerValAlaValThrGlnGlnLeuTyrAlaGln-----ProAsnGlu 29
Db 214 CTACTATCGATTGGTATTTCAGCGCGAGCTTTTGCACAGCGGACACGTCGCCAGAT 273
Qy 30 SerLeuProThrValGluLeuGluProValValIleThrIleAspLysSerGlyMetala 49
Db 274 TCGCTACGAGTACAAATCTTCAGACCGTCACGGTCTATTCTTACACGACGCGCGTACCT 333
Qy 50 LeuAlaAsnArgIleThrGlnMetProHisThrLysValIleTyrGluGluGlnIle 69
Db 334 CTG-----AAAAAGATACCGGCAAGTGGAACTCATCTCATCGCGCAACATC 381
Qy 70 GlnGluGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSer 89
Db 382 AAGCAG-----TCGCGCTTTTAAACAAATGACCGCATCTCTC---AAGACGCAAAAGTTTCG 432
Qy 90 LeuGlyValSerSer-----GlyThrThrSerAsnPhe-----GlyGlnThrMet 104
Db 433 CTCGATGTCAATACATACCCGGGCTTTAGTTCGAACATCGGTATCCGCGGTTCGAAGCCC 492
Qy 105 HisGlyArgGlnValGlnPheLeuLeuAsnGlyValProLeuThrGlySerArgAspIle 124
Db 493 TCCGGCAAGTATGTAACCGTATTGGTAAACGGCATCCCT---GCCGGAAACGGACAAATATC 549

```

QY 125 SerArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGly 144  
 DB 550 TCT-----ACGCTCAACACGACCAACATCGAACCAATCGATCTCTCAAGAGC 597  
 QY 145 Ala-----ThrSerIleTyrGlySerGlyAlaThrGlyGluLeuIleAsnIleValThrLys 163  
 DB 598 CCGTCTCTTCCATCTACGACCAATGCCATGGCGGTGGTGAACATCATCACCAC 657  
 QY 164 SerAspLeuGluGluGluInPheGluThrArgIleGlyValIleGlySerLysLeuSer 183  
 DB 658 AATCCAGAGACAG-----ATCATGGCAAC---GTTTCT 690  
 QY 184 SerGluGlyIleGlyTyrGlnValGlyGlnSerValAlaGlyValSerGluAsnGly 203  
 DB 691 CTCTTCGGCGAGTAC-----CAGACCATGGCCGATCATCACTGGGTGGC 741  
 QY 204 -----ValLeuAlaArgLeuAspValAsp 211  
 DB 742 CGCTTCGAGGAAATTTCTCATCTGATCTTACGTGGCTTGACAGCAAGCAAGAGAC 801  
 QY 212 TyrArgThrThrGlyGlyAlaPheAspAlaAsnGlyLysArgIleAlaProGluProAla 231  
 DB 802 TATTAAGACCGGATCAACAACTTCTCATCTGAGCAAACTG-----GAAAGAGCT 852  
 QY 232 GlnThrAspLysGlnAspSerLysSer----- 240  
 DB 853 ATAGTAGATGTAATGCTACCAAAAACAGAAATGAAGGAGCGACTACTACTAGCA 912  
 QY 241 -----LeuSerValAsnThrAsnValAspTyrGlnLeuAspLysGln 255  
 DB 913 ACGGAGCGTTCGGTTTCGGTATCGACTTCACGCCGAAATGGTGGTGAATCTGTATCA 972  
 QY 256 AsnIleAsnLeuAlaLeuThrIleTyrAsnAspLysGlnAspTyrAlaProAsp 275  
 DB 973 AACGTATCTCTCGGA-----CATCGATCCCGTA 1002  
 QY 276 TyrGlyAsnArgLeuAlaValLeuPheGlyGluLysProSerLeuAsnAlaIleLysGly 295  
 DB 1003 GGAGGATCTATATGGGGCGGTTTACGAGAAATCAAAAATCTGAAT----- 1050  
 QY 296 LeuSerLeuSerGluGlnProLysThrThrLysSerThrPheAsnIleAsnTyrHisHis 315  
 DB 1051 -----CGTCTTCACCTCTTTCAGCTGCTCGCAACAT 1086  
 QY 316 AspAspLeuTyrGlyAsnThrIleAsnThrAsnAlaTyrTyrArgArgGluLysGlyAsp 335  
 DB 1087 GGC-----TGCCACACGCTTCAATCTCCCTACTTCAACATAGAAATCGAG 1137  
 QY 336 PheTyrProPheValAlaProPheSerIleAlaLysAlaLeuProIleLeuGlnSerMet 355  
 DB 1138 AACTATTAACAATGCCGATCCGCGTTTC-----ATC 1170  
 QY 356 AsnLeuProSerAlaThrLeuAspAlaTyrThrLysAlaProGlnAlaArgAlaTyrGly 375  
 DB 1171 AACTACAAAGC-----GACTACTACACC-----TATGGT 1200  
 QY 376 ValLeu 377  
 DB 1201 GCCCTA 1206  
 RESULT 10  
 AAF26439  
 ID AAF26439 standard; DNA; 885 BP.  
 XX  
 AC AAF26439;  
 XX  
 DT 02-MAY-2001 (first entry)  
 XX  
 DE Pseudomonas sp heavy metal transporter encoding DNA ORF11136.  
 XX  
 KM Heavy metal transporter; iron transporter; transgenic plant;  
 KW homeostasis regulator; heavy metal ion; trace element; soil pollution;  
 plant growth promoter; plant development; ds.

XX  
 OS Pseudomonas sp.  
 XX DE19934720-A1.  
 PN  
 XX 25-JAN-2001.  
 PD  
 XX  
 PF 23-JUL-1999; 99DE-1034720.  
 XX  
 PR 23-JUL-1999; 99DE-1034720.  
 XX  
 PA (TIGR-) TIGR INST GENOMIC RES.  
 PA (QUTA-) QUITAGEN GMBH.  
 PA (GFRB-) GFS BIOTECHNOLOGISCHE FORSCHUNG MBH.  
 PA (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.  
 PA (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.  
 XX  
 DR WPI; 2001-160596/17.  
 XX  
 PT New DNA encoding iron or heavy metal transporters, useful for promoting  
 XX growth of plants on polluted soil -  
 PT  
 PS Claim 3a; Page 27; 54pp; German.  
 XX  
 CC This invention describes novel DNA sequences (1) that encode iron or  
 CC heavy metal transporters (11). The invention also describes (1) a  
 CC recombinant expression vector (11) containing (1); (2) prokaryotic or  
 CC eukaryotic cells (11) transformed with (1) or (11); (3) production of  
 CC (11) by culturing (11); (4) (partial) expression products (11) of (1)  
 CC and synthetic proteins or peptides with the same sequences; (5)  
 CC antibodies (Ab) specific for (11); (6) hybridoma cells that produce  
 CC monoclonal Ab; and (7) transgenic plants that contain (11). The iron or  
 CC heavy metal transporters encoded by (1) transport heavy metal ions across  
 CC cell walls and regulate homeostasis of trace elements (1), and their  
 CC fragments are useful for: (1) expression of (1); (2) as probes and  
 CC primers for detection, isolation and amplification of full length cDNA  
 CC sequences; and (3) producing transgenic plants. (11) are used to promote  
 CC growth, development and yield of plants, particularly Leguminosae,  
 CC especially when growing in soil polluted by heavy metals injurious to  
 CC plants. They also improve homeostasis of iron and trace elements. Host  
 CC cells that express the iron or heavy metal transporters take up heavy  
 CC metals, so reduce pollution of soil and release iron and other trace  
 CC metals which improves soil quality and protect plants against pollutants  
 CC from the soil.  
 XX  
 SQ Sequence 885 BP; 192 A; 302 C; 248 G; 141 T; 2 other;  
 XX  
 Alignment Scores:  
 Pred. No.: 5.09e-08 Length: 885  
 Score: 180.00 Matches: 72  
 Percent Similarity: 45.83% Conservative: 60  
 Best Local Similarity: 25.00% Mismatches: 116  
 Query Match: 8.79% Indels: 40  
 DB: 22 Gaps: 13  
 US-09-889-746-2 (1-400) x AAF26439 (1-885)  
 QY 54 IletHrGlnMetProHisThrThrLysValIleTyrGluGluGlnIleGlnGluAla 73  
 DB 19 CTGCGGATATCCCGCAACCGTCAACGCTTCCCGAAAGCGTATCAAGACCA--- 75  
 QY 74 ThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSer 93  
 DB 76 ---GGCGACAGTCCATGAAAGAGCTGCTGAAGAAC---GTCCCGCATCGGCTTCTC 129  
 QY 94 SerGlyThrThrSerAsnPheGlyGlnThrMetHisGlyArgGlnValGlnPheLeu 113  
 DB 130 AACGGGACGGCAAGCGCATCAGGAGCAATCCGCTTCAAGGCC-----ATT 180  
 QY 114 AsnGlyValProLeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAsnPro 133  
 DB 181 GGCAATGTATCATAGACGGCGTGGTGAACGACCGCTGACTTTCGGCATCTG----- 234

```
QY 134 AsnGlnValAlaArgIleGluValLeuSerGly---AlaThrSerIleTyrGlySerGly 152
DB 235 TCCAAACATCGAGCGTGTCCAGGTGATCAAGGGCCCGCGGTCTGTATGCCGTGGC 294
QY 153 AlaThrGlyGlyLeuIleAsnIleValThrLys-----SerAspLeuGluGluGln 170
DB 295 TCCTCGGGGGCTGATCAACAGCATCAGCAAAACCAAGCTTCACGCCCAAGCGCGAA 354
QY 171 PheGluThrArgIleGlyValHisGlySerLysLeuSerSerGluGlyIleGlyTyrGln 190
DB 355 GTCGGCATGAACGTCGACAGCAAGCAACCGCAACCCAGTTCGACACCGCTGGGCC 414
QY 191 ValGlyGlnSerValAlaGlyValSerGluAsnGlyAsnValLeuAlaArgLeuAspVal 210
DB 415 GACCCGCAAGC-----AACCCAG 432
QY 211 AspTyrArgThrThrGlyGlyAlaPheAspAlaAsnGlyLysArgIleAlaProGluPro 230
DB 433 GCTACCGGTCACCGTCCCTCGAAGACAGGACACCTTCGCG-----GATGAT 483
QY 231 AlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsnThrAsnValAspTyrGln 250
DB 484 GGCTACATCGACCGCAAG-----GCCATCGCACCGTCCGCTACTTCGCT 528
QY 251 LeuAspAspLysGlnAsnIleAsnLeuAlaLeuThrHisTyrAsnAspLysGlnAspThr 270
DB 529 CTCCTCGGACGATCTGGAACCTGGAACCTGGCGCCACCTACTCTGACACAAAGCGCTGATC 588
QY 271 AspTyrAla---ProAspTyrGlyAsnArgLeuAlaValLeuPheGlyGlyLys----- 287
DB 589 GACTTCGGCATCCCGCCCTGGGCAACCCCGGTGGAGCTGGACCGCGACAAAGCGCTTT 648
QY 288 ProSerLeuAsnAlaIleLysGlyLeuSerLeuSerGluGlnProLysThrThrLysSer 307
DB 649 GGCCTCTGGCGATGCTGACCAAGATTACGCCCGCAGGAGGTGTTCTCGCTTACAGCCAGC 708
QY 308 ThrPheAsnIleAsnTyrHisAspAspLeuTyrGlyAsnThrIleAsnThrAsnAla 327
DB 709 -----CTGGACTACCGCATCAACGACAACTTCACCCCTGACC---AACACCAAGCGCT 756
QY 328 TyrTyrArgArgGlyGlyArg 335
DB 757 TACTACCGTACGACTTGGACCGC 780
RESULT 11
ID AAX91566
XX AAX91566;
AC AAX91566;
XX
XX
XX 25-AUG-1999 (first entry)
XX
XX Porphorymonas gingivalis protein PG13 ORF encoding DNA.
XX
XX Porphorymonas gingivalis; PG; periodontal disease; gingivitis;
XX vaccine; antigenic; ds.
XX
XX Porphorymonas gingivalis.
XX
XX WO929870-A1.
XX
XX 17-JUN-1999.
XX
XX 10-DEC-1998; 98AU-0001023.
XX
XX 04-AUG-1998; 98AU-0005028.
XX
XX 10-DEC-1997; 97AU-0000839.
XX
XX 31-DEC-1997; 97AU-0001182.
XX
XX 30-JAN-1998; 98AU-0001546.
XX
XX 10-MAR-1998; 98AU-0002264.
XX
XX 09-APR-1998; 98AU-0002911.
XX
XX 23-APR-1998; 98AU-0003128.
XX
XX 05-MAY-1998; 98AU-0003338.
```

```
PR 22-MAY-1998; 98AU-0003654.
PR 29-JUL-1998; 98AU-0004917.
XX (CSLC-) CSL LTD.
XX
XX Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX
XX WPI; 1999-385613/32.
DR P-PSDB; AAY34348.
XX
XX Antigenic Porphorymonas gingivalis peptides for preventing
PT gingivitis
XX
XX Claim 12; Page 115; 588pp; English.
XX
XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphorymonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphorymonas gingivalis. Probes can
CC be used to detect Porphorymonas gingivalis in standard hybridisation
CC assays. Porphorymonas gingivalis is involved in periodontal disease
CC especially gingivitis.
XX
XX Sequence 2289 BP; 588 A; 563 C; 553 G; 585 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 2,59e-07 Length: 2289
Score: 178.50 Matches: 83
Percent Similarity: 41.99% Conservative: 56
Best Local Similarity: 25.08% Mismatches: 147
Query Match: 8.72% Indels: 45
DB: 20 Gaps: 16
US-09-889-746-2 (1-400) x AAX91566 (1-2289)
QY 31 LeuProThrValGluLeuGluProValValIleThrIleAspLysSerGlyMetAlaLeu 50
DB 334 CTGCGAAGCAACACTTGGAGGAGTCTGCTTACC-----GGTACCGGTACA 381
QY 51 AlaAsnArgIleThrGlnMetProHisThrLysValIleTyrGluGluGlnIleGln 70
DB 382 CGTTACCGCTTGGTCGATGCTCTCTGCGCAACGGAAGTCTTACCGCTAAGGACATA--- 438
QY 71 GluGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGln---LeuIleProSer 89
DB 439 -----GCCTCTTCTCGGCTCCTACTTCGAGGCGCTTATTCAGGCGCTGAGTCCGTCT 492
QY 90 LeuGlyValSerSerGlyThrThrSerAsnPheGlyGlnThr-----MethIleGlyArg 107
DB 493 TTGTGCTTCGCGCCCAATCTGATGGGCTCTTTCATGCACTGMAACGGCTTAGCAGTAAG 552
QY 108 GlnValGlnPheLeuLeuAsnGlyValProLeuThrGlySerArgAspIleSerArgGln 127
DB 553 TATATCCTCATCTTATCGATGTTAAGCGTGTGACGCGATGTAGCGCGTCAGGCCGAT 612
QY 128 LeuAsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyAla---Thr 146
DB 613 TTGAGTCGTATTTCTCTGATCAGATCGAAGGATCGAACTGGTGAAGAGGTGCTTCGAGT 672
QY 147 SerIleTyrGlySerGlyAlaThrGlyLeuIleAsnIleValThrLysSerAspLeu 166
DB 673 TCGCTCTACGATCGATGCCATCGCGGGGTAACTCAATGTGATCACAAGAAAGTAACG 732
QY 167 GluGluGluGlnPheGluThr-----ArgIleGlyValHisGlySerLysLeuSerSer 184
DB 733 AATCGACTGAGTGCATATACGTCATCGCATATCGAAGTACAACGATCGGCAACCAAT 792
QY 185 GluGlyIleGlyTyrGlnValGlyGlnSerValAlaGlyValSerGluAsnGlyAsnVal 204
DB 793 ACTTCGCTCGATATAAACATCGGTAAAG-----TTCAGTAGCAATACCACTAT 840
```

QY 205 LeuAlaIArgLeuAspValAspTyrArgThrThrGly-----GlyAlaPheAspAla 221  
 DB 841 TTCTTC-----TACCATAGCGATGGCTGGCAGAAATAGTCGTTGAAATA 885  
 QY 222 AsnGlyValArgIleAlaProGluProAla-----GlnThrAspLysGlnAspSerLys 239  
 DB 886 AAAAGAAAAGAGATCCGGCGGACCGCTTGGAGAAAACGATATAGAAAACCTTTTCGT 945  
 QY 240 SerLeuSerValAsnThrAsnValAspTyrGlnLeuAspLysGlnAsnIleAsnLeu 259  
 DB 946 GCA---CAGAAAATCAGGGGTGAGCAATCCGTTTCTTATATGCACTAAACAATCTT 1002  
 QY 260 AlaLeuThr-----HisTyrAsnAspLysGlnAspThrAspTyrAlaProAspTyr 276  
 DB 1003 AGCTTACGGCGAATGACGAGTACCAATAAACGTACG-----ATCTTCACTCCGACCTTT 1056  
 QY 277 GlyAsnArgLeuAlaValLeuPheGlyGlnLysProSerLeuAsnAlaIleLysGlyLeu 296  
 DB 1057 TCCGAAAAGAGCCTATGACATGATTAATCGT---GCTTGAACGCGCTTCACTCGGTACG 1113  
 QY 297 -----SerLeuSerGluGlnProLysThrThrLysSer 307  
 DB 1114 AACTATCTTTTCCCAATGATGCTGCATACGCTTCTTTCATGCGCTTACATCGCTTC 1173  
 QY 308 ThrPheAsnIleAsnTyrHisHisAspAspLeuTyrGlyAsnThrIleAsnThrAsnAla 327  
 DB 1174 CGTTGGATTTTGTATCATCATGACAGACAGACAGTACGAGCCGTGATCAACACCAAGT 1233  
 QY 328 TyrTyrArgArgGluLysGlyArgPheTyrPro 338  
 DB 1234 -----CAGACCGAGCAACCCACATCTCTCCG 1260

RESULT 12  
 AAS82047  
 ID AAS82047 standard; cDNA, 1827 BP.

AC AAS82047;  
 XX  
 DT 13-FEB-2002 (first entry)  
 DE DNA encoding novel human diagnostic protein #17851.  
 XX  
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 OS Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 PD 11-OCT-2001.  
 PF 30-MAR-2001; 2001WO-US08631.  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 XX  
 PA (HYSR-) HYSER INC.  
 PI Drmanac RT, Liu C, Tang YT;  
 DR MPI; 2001-639362/73.  
 DR P-PDB; ABG17860.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX  
 PS Claim 1; SEQ ID No 17851; 103bp; English.  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.int/pub/published\_pct\_sequences.  
 XX

SO Sequence 1827 BP; 453 A; 474 C; 462 G; 438 T; 0 other;

Alignment Scores:  
 Pred. No.: 9.05e-07 Length: 1827  
 Score: 171.50 Matches: 92  
 Percent Similarity: 34.99% Conservative: 49  
 Best Local Similarity: 22.83% Mismatches: 141  
 Query Match: 8.38% Indels: 121  
 DB: 23 Gaps: 19

US-09-889-746-2 (1-400) x AAS82047 (1-1827)

QY 17 ValAlaValThrGlnGlnLeuTyrAlaGlnProAsnGlnSerLeuProThrValGluLeu 36  
 DB 658 ATGTCTGTTTATGACAGCGCGGTTGAACGAAAGAAAGACATTCACCGTTACCGCT 717  
 QY 37 GluPro-----ValValIleThrIleAsp 44  
 DB 718 GCACCTGGCGCGCAAGAAAGCGCATGGGGCGCTGTCGAACATTTGCGGCGACAGTCT 777  
 QY 45 LysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrThrLysValIle 64  
 DB 778 GCTAACCGGCACTAAACCGATACCGCATTAACAAAGTCCACAGCTCTATTCTGTTGTC 837  
 QY 65 TyrGlnGlnGlnIleGlnGlnGlnAlaThrGlySerArgGlnLeuAlaAspValMetAla 84  
 DB 838 ACCGCGAAGAGATG-----GCGCTGCATCAGCGCAAGTCCGTAAGAA 882  
 QY 85 GlnLeuIleProSerLeuGlyValSerSerGlyThr-----ThrSerAsnPhe----- 100  
 DB 883 GCGCTTAGCTACACGCGCGGTGTCTCTGTGGTACCGCGGCACTCAACACTATGAC 942  
 QY 101 -----GlyGlnThrMetHisGlyArgGlnValGlnPheLeuLeuAngly 115  
 DB 943 CACCTGATCATTCGCGCTTTGGCGGACGAGCAAGCCAGCAATATCATCTGAATGCG 1002  
 QY 116 ValProLeuThrGlySer-----ArgAspIleSerArgGlnLeuAsnSerIleAsnPro 133  
 DB 1003 CTGAAGTTGACGGCAACTTCTATAACGATGCG-----GTCAATGACCCG 1047  
 QY 134 AsnGlnValAlaArgIleGlnValLeuSerGlyAlaThrSerIle---TyrGlySerGly 152  
 DB 1048 TATATGCTGGAACGCGCTGAATATATGCGTGGCGCGCTTCCGTTCATTAAGTAAAGC 1107  
 QY 153 AlaThrGlyGlyLeuIleAsnIleValThrLys-----SerAspLeuGlnGlu 168  
 DB 1108 AATCTGCGGCGCTGTGATATGATGTCACAGACGCTCCACACCAACCGCTGAAGAA 1167  
 QY 169 GluGlnPheGlnThrArgIleGlyValHisGlySerTyrLeuSerSerGlyGlyIleGly 188  
 DB 1168 GTTCAGTTT-----AAAGCCGCTACTACACACCTG--- 1197  
 QY 189 TyrGlnValGlyGlnSerValAlaGlyValSerGluAsnGlyAsnValLeuAlaArgLeu 208

```

Db 1198 TTCCAGACTGGTTTGGACTTTAGCATTCGTTGGGATGATGACGGTGTACTCT----- 1251
Qy 209 AspValAspTyrArgThrThrGlyGlyAlaPheAspAlaAsnGlyLysArgIleAlaPro 228
Db 1252 -----TATCGCTGACCGGCTTGGCGCTTCTGCCAAT----- 1284
Qy 229 GluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsnThrAsnValAsp 248
Db 1285 -----GCCACGACAAAGGTCAGAAAGCAGCAGCGTATGCTATTATTCACCGCGTTCACC 1338
Qy 249 TrpGlnLeuAspAspLysGlnAsnIleAsnLeuAlaLeuThrHisTyrAsnAspLysGln 268
Db 1339 TGGCGTCGGATGATAAACCAATTTTACCTTC-----CTTCTTACTTCCAGACGAGCGG 1395
Qy 269 AspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeuPheGlyGluLysPro 288
Db 1396 GATACCGGTTAT-----TACGGCTGGTTGCGG 1422
Qy 289 SerLeuAsnAlaIleLysGlyLeuSerLeuSerGluGlnProLysThrThrLysSerThr 308
Db 1423 AAAGAGGGAACCGTTGAGCGGCTG-----CCGAACGGTTAAGCGTTTGGCGACAGAC 1473
Qy 309 PheAsnIleAsnTyrHisHisAspAspLeuTyrGlyAsnThrIleAsnThrAsnAlaTyr 328
Db 1474 TTTAAT-----GATAGGCGAAGAACACGCCCTAT 1503
Qy 329 TyrArgAlaGluLysGlyArgPheTyrProPheValAlaProPheSerIleAlaLysAla 348
Db 1504 TCTCGTATGAGAG----- 1518
Qy 349 LeuProIleLeuGlnSerMetAsnLeuProSerAlaThrLeuAspAlaTyrThrLysAla 368
Db 1519 -----ATGAAC-----GAA 1527
Qy 369 ProGlnAlaArgAlaTyrGlyValLeuGlnSerGlnSerLysAlaGluValLeuGlyArg 388
Db 1528 CCGAAACCGGTATACCGCTGTGTCGGAAGACGGAACCGTTGAAACCGCTG----- 1581
Qy 389 ValProAsn 391
Db 1582 ---CCGAC 1587
RESULT 13
AAF26420
ID AAF26420 standard; DNA; 2127 BP.
XX
XX
XX AAF26420;
XX
XX 02-MAY-2001 (first entry)
XX
XX Pseudomonas sp heavy metal transporter encoding DNA ORF03875.
XX
XX Heavy metal transporter; iron transporter; transgenic plant;
XX homeostasis regulator; heavy metal ion; trace element; soil pollution;
XX plant growth promoter; plant development; ds.
XX
XX Pseudomonas sp.
XX
XX DE19934720-A1.
XX
XX 25-JAN-2001.
XX
XX 23-JUL-1999; 99DE-1034720.
XX
XX 23-JUL-1999; 99DE-1034720.
XX
XX (TIGR-) TIGR INST GENOMIC RES.
XX (QUIA-) QUIAGEN GMBH.
XX (GBF-) GBS BIOTECHNOLOGISCHE FORSCHUNG MBH.
XX (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
XX (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.

```

```

DR WPI; 2001-160596/17.
XX
XX New DNA encoding iron or heavy metal transporters, useful for promoting
XX growth of plants on polluted soil.
XX
XX Claim 3a; Page 15-16; 54pp; German.
XX
XX This invention describes novel DNA sequences (I) that encode iron or
XX heavy metal transporters (II). The invention also describes (1) a
XX recombinant expression vector (III) containing (I); (2) prokaryotic or
XX eukaryotic cells (III) transformed with (I) or (III); (3) production of
XX (II) by culturing (III); (4) (partial) expression products (IV) of (I)
XX and synthetic proteins or peptides with the same sequences; (5)
XX antibodies (Ab) specific for (IV); (6) hybridoma cells that produce
XX monoclonal Ab; and (7) transgenic plants that contain (III). The iron or
XX heavy metal transporters encoded by (I) transport heavy metal ions across
XX cell walls and regulate homeostasis of trace elements. (I), and their
XX fragments are useful for: (1) expression of (II); (2) as probes and
XX primers for detection, isolation and amplification of full length cDNA
XX sequences; and (3) producing transgenic plants. (III) are used to promote
XX growth, development and yield of plants, particularly Leguminosae,
XX especially when growing in soil polluted by heavy metals injurious to
XX plants. They also improve homeostasis of iron and trace elements. Host
XX cells that express the iron or heavy metal transporters take up heavy
XX metals, so reduce pollution of soil and release iron and other trace
XX metals which improves soil quality and protect plants against pollutants
XX from the soil.
XX
XX Sequence 2127 BP; 458 A; 710 C; 611 G; 346 T; 2 other;

Alignment Scores:
Pred. No.: 7.63e-05 Length: 2127
Score: 152.50 Matches: 84
Percent Similarity: 38.93% Conservative: 62
Best Local Similarity: 22.40% Mismatches: 122
Query Match: 7.45% Indels: 107
DB: 22 Gaps: 19

US-09-889-746-2 (1-400) x AAF26420 (1-2127)
Qy 9 TrpLeuSerLeuProLeuLeuSerValAlaValThrGlnGlnLeuTyrAlaGlnProAsn 28
Db 13 TTGCTTCGCTTTGTGCTTCAT---GCTGTCTCCCGCTGGCGCTTTGCCGAGCT--- 66
Qy 29 GluSerLeuPro-----ThrValGluLeuGluProValAlaIleThr----- 42
Db 67 GAAACGCTCAGCGACCCCTAACAGATCGAACTCCAGGCGCTGAGCATCACCAGTACTGCC 126
Qy 43 -----IleAspLysSerGlyMetAlaLeuAla 51
Db 127 GACAGCGAAGCGCGCGCGCGCTCGAAGGCTACAGGCGCACCCGCTCGCGCAGTGCC 186
Qy 52 AsnArg-----IleThrGlnMetProHisThrThrLysValIleTyrGluGlu 67
Db 187 ACCGCGACCGACGCGCTCGACGAAACCCCGCAATCGGTCAGCGTGGTCCCAACGAT 246
Qy 68 GlnIleGlnGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIle 87
Db 247 GTGCTGGTAGAC-----ACGGGCGCCACGCGCTCGCAGGAAGGCGCTG----- 288
Qy 88 ProSerLeuGlyValSerSerGlyThrThrSerAsnPheGlyGln----- 102
Db 289 ---GACTACGCGCGTGGTGTGCGCGCGCGCAACAACCTTCGCTGGCGCAGGCGCTGACAACC 345
Qy 103 ---ThrMetHisGlyArgGlnVal---GlnPheLeuLeuAsnGlyValProLeuThrGly 120
Db 346 TTCACGTGCGTGGCTTCACGACCGCGCGAGTCTATCCCAACGCGCTTTCCGATCAACCGC 405
Qy 121 SerArgAspIleSerArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGlu 140
Db 406 GCGTACCC-----AACGCGCGGATGCCAACACCGCTCGAAGCAGTGGAG 450
Qy 141 ValLeuSerGly---AlaThrSerIleTyrGlySerGlyAlaThrGlyGlyLeuIleAsn 159

```

```

Db 451 GTCATCGTGGCCCGGCGACAGCTGTACGCGCGTGGCCACCCCGCGGTTCAC 510
QY 160 ILVAlThrIysSerAspLeuGluGluGluPheGluThrArgIleGlyValHisGly 179
Db 511 GTGTGACGACGAGCGCGTGTCCGAGTCCAGGTCACTCCCTT-----GGC 555
QY 180 SerIysLeuSerSerGluGlyIleGlyTyrGlnValGlyIleSerValAlaGlyValSer 199
Db 556 AGCCAGTTCACGACGAGCGGCGATGACCGGTCCACCTTCATGTCACCGGCGCGTACG 615
QY 200 GluAsnGlyAsnValIleuAlaArgLeuAspValAspTyrArgThrGlyGlyAlaPhe 219
Db 616 CAGACGCGCTCGCTGCGCTACCGCTGACGTA---CTGGGTGAGGCGCGACAGCTTC 672
QY 220 AspAlaAsnGlyLysArgIleAlaProGluProAlaGlnThrAspLysGluAspSerLys 239
Db 673 -----CGCGACGATTCGAAAGCGAA 693
QY 240 SerLeuSerValAsnThrAsnValAspTyrGlnLeuAspLysGluAsnIleAsnLeu 259
Db 694 CGCTACGATGTGCGCCCTGTATCAGCTGAGGTCAACGACGCGACGAAAGTCACTTC 753
QY 260 -----AlaLeuThrHisTyr----- 264
Db 754 GAAGCGCATTCATGCGCAACAACCCCTCGATGCGCGCTGACCCCTACCCAC 813
QY 265 -----AsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsn----- 278
Db 814 CAGACTGGCAGCGCTCGCGGATCC---TACATTTGGAGAAAGGCGACGACACCTG 870
QY 279 -----ArgLeuAlaValLeuPheGlyGlyLysPro 288
Db 871 CTGCACACGACACAAACATGATGCGCCAGTGGCTTGAGACACCTGCTCAACGACACTGG 930
QY 289 SerLeu-----AsnAlaIleLysGly 295
Db 931 ACCTGTGGCGGTGATTCAGTACCTCGATGCTGCTCAAGGAGTATGCGGTGAGGGG 990
QY 296 LeuSerLeuSerGluGlnProLysThrThrLysSerThrPheAsn 310
Db 991 AGCATTTACTGGCAGACGCGCGACCTGTGACGGAATTCATC 1035

RESULT 14
ID AAA59216 standard; DNA; 2169 BP.
XX AAA59216;
AC AAA59216;
XX 07-NOV-2000 (first entry)
DT
XX
DE DNA encoding a Neisseria meningitidis BASB053 polypeptide.
XX
KM BASB053; Neisseria meningitidis infection; vaccine; ss.
XX
OS Neisseria meningitidis.
XX
FH Key Location/Qualifiers
FT CDS 1..2169
FT /tag= a
FT /product= "BASB053"
XX
XX MO200042193-Al.
XX
XX 20-JUL-2000.
XX
XX 10-JAN-2000; 2000MO-EP00137.
XX
XX 15-JAN-1999; 99GB-0000959.
XX
XX 28-JAN-1999; 99GB-0001903.
XX
XX (SMIT ) SMITHKLINE BEECHAM BIOLOGICALS.
XX

```

```

PI Ruelle J;
XX
XX MPI; 2000-476062/41.
DR P-PSDB; AAB07697.
XX
XX New Neisseria meningitidis polypeptide useful for diagnosis of
PT Neisseria infection and for development of vaccines against such
PT infection -
XX
XX Claim 11; Page 55; 92pp; English.
XX
XX The present sequence encodes a Neisseria meningitidis BASB053
CC polypeptide. The BASB053 polypeptide, or an antibody immunospecific
CC for BASB053 may be identified in a biological sample in order to
CC diagnose a Neisseria meningitidis infection in an animal. The BASB053
CC polypeptides and polynucleotides may be used as vaccines, for
CC generating an immune response in an animal. A composition comprising
CC at least one antibody immunospecific for BASB053 may be used to
CC treat humans infected with Neisseria meningitidis.
XX
SQ Sequence 2169 BP; 540 A; 663 C; 568 G; 398 T; 0 other;

Alignment Scores:
Pred. No.: 8.76e-05 Length: 2169
Score: 152.00 Matches: 102
Percent Similarity: 37.72% Conservative: 70
Best Local Similarity: 22.37% Mismatches: 159
Query Match: 7.43% Indels: 125
DB: 21 Gaps: 24

US-09-889-746-2 (1-400) x AAA59216 (1-2169)
QY 18 AlAlValThrGlnGlnLeuIleuValAsnGluSerLeuProThrValGluLeuGlu 37
Db 58 GCATTCCTGCTTGGCTTTTCCGCAACAACGCGGATTTGAAACCGTCCACATCAA 117
QY 38 ProVal-----ValIleThrIleAspLysSerGlyMetAlaLeuAlaAsnArg 53
Db 118 GCGCAGCGTTGTAACAACGCGATTTGACCGAATAACGCGCATTAACGCTGTTGCC 177
QY 54 IleThr-----GlnMetProHisThrThrLysVal 63
Db 178 GTCACCGTGGCACAATAATCCCGCTTGTGGCGGAATTCGCAATCCGTCAGTATC 237
QY 64 IleTyrGluGluGlnIleGlnGlnAlaThrGlySerArgGlnLeuAlaAspValMet 83
Db 238 ATCACAACCGACGAGGTCAAGCGCAATGTTGATAC-----TTTGACCAAGTTG 288
QY 84 AlaGlnLeuIleProSerLeuGlyValSerSerGlyThrThrSerAsnPheGlyGlnThr 103
Db 289 GCGCGCAAAACGCGCGCTGCGCGGTGAGCAAGATGACGAGGCTTGGTTAC 348
QY 104 MetHisGlyArgGlnVal---GlnPheLeuLeuAsnGlyValProLeuThrGlySerArg 122
Db 349 GCGCGCGGTTACGAATACAGCGCAATACATCAGCGCGCTGCC----- 393
QY 123 AspIleSerArgGlnLeuAsnSerIleAsn-----ProAsnGlnValAla 137
Db 394 -----CGCGAGATGACGAGTATCAACGCGACCGCTGCCAATGTTGGCTTGCAC 444
QY 138 ArgIleGluValLeuSerGlyAlaThrSerIleTyr---GlySerGlyAlaThrGlyGly 156
Db 445 CCGGTGAGATGATGCGCGCGCGCGGACGACTGTTGACACGACGCGCGAGATGGCGGT 504
QY 157 LeuIleAsnIleValThrLysSerAspLeuGlu----- 167
Db 505 ATCGTGAATCTGTGGCAACCGCCGACCAAGCGCTTCCAAAGTATGCTGCGGAGGG 564
QY 168 -----GluGlnGlnPheGluThrArgIleGlyValHisGlySerLysLeuSerSer 184
Db 565 TTGCTAGTACGACAAACAAATATTAAGCGGAGCGGAGCTATCGGCGACG---CTCAATTCA 621
QY 185 GluGly-----IleGlyTyrGlnValGlyGlnSerValAlaGlyVal 198

```

```

Db 622 GACGGCAGCGTGGCGCGCGCTGATGGCGCAGACCGTGGCGCGCTCTCCGGTCCCGCC 681
Qy 199 SerGluAsnGlyAsn-----ValLeuAlaArgLeuAspValAspTyrArg 213
Db 682 GAGAAACACACCGCGCAGCAACCTTCTACGGCGCGCGGATGGACATCAACCCCGAT 741
Qy 214 ThrThr-----GlyGlyAlaPheAspAlaAsnGlyLysArgIleAlaProGlu----- 229
Db 742 ACGGTTTGGCGCGCGGCTATCTTACCAGCAACGCCACCTCGCGCGCTACAAACGGCTTG 801
Qy 230 ProAlaGln-----ThrAspGlyGlnAspSerLysSerLeuSerValAsnThrAsnValAsp 248
Db 802 CCAGCGGATGCAATACAAATATACCGTCCCTCGCGCAACACGATTTGTGCGCGCGGAT 861
Qy 249 TrpGln-----LeuAspAspLysGlnAsnIleAsnLeuAlaLeuThrHis----- 263
Db 862 TGGACAAATTTAAATAGACACCGCAGCGTGTGGCGATTTGAAACATTACTTTCGCG 921
Qy 264 -----TyrAsnAspLysGln-----AspThrAspTyr 272
Db 922 AACGGCGGCTACGGCAAGTCGGTATCGCGTATTCGACCGCGATCGCGACTCCCAACTAT 981
Qy 273 Ala-----ProAspTyrGlyAsnArgLeuAlaValLe 283
Db 982 GCCTTTGGCGCAGCAAGCTGGCGATGAAACCCCGCGCAGCGCGCGGCTGCAATACG 1041
Qy 283 uPheGlyGluLysPro-----ProAspTyrGlyAsnArgLeuAlaValLe 288
Db 1042 GCTGACGCAAAAGCTCGCGGTGGTGGTACAGAAATCAACAAACCCCTCGCG 1101
Qy 289 -----SerLeuAsnAlaIleLysGlyLeuSerLeuSerGln 300
Db 1102 TTGACGGCGCTACAGCAGCGCTTTCGCGTATTCGCGCAATACGCCCAAGATTTGTCATC 1161
Qy 300 uGlnProLysThrThrLysSerThrPheAsnIleAsn-----TyrHis 315
Db 1162 GCGCGCGATTAACACCGCTTTCGCGACCAACGACCAAGCGCGTACTACTTTATATGCA 1221
Qy 315 eAspAspLeuTrpGlyAsnThrIleAsnThrAsnAlaTyrTyrArgGluLysGlyArg 335
Db 1222 CCGCGCGCGCTGGCTTTAAACG-----AGTTCCGACGATACCGCAGG----- 1264
Qy 335 gPheTyrProPheAlaAlaProPheSerIleAlaLysAlaLeuProIleLeuGlnSerMe 355
Db 1265 -----TTGATTGATTGCCAACGCGCGCAAGCGGTGCGCGGTACAGCCATACC 1314
Qy 355 tAsnLeuPro-----SerAlaThrLeuAspAlaTyrThrLysAlaPro 369
Db 1315 GTCGTACGAAACCTCGACGAATTCGGCATTTACGGCAAAATCCA 1360
RESULT 15
ID AAF26411 standard; DNA; 2586 BP.
AC AAF26411;
XX
DT 02-MAY-2001 (first entry)
XX
DE Pseudomonas sp heavy metal transporter encoding DNA ORF09606.
XX
KW Heavy metal transporter; iron transporter; transgenic plant;
KW homeostasis regulator; heavy metal ion; trace element; soil pollution;
KW plant growth promoter; plant development; ds.
XX
OS Pseudomonas sp.
XX
FN DE19934720-A1.
XX
PD 25-JAN-2001.
XX
PF 23-JUL-1999; 99DE-1034720.
XX

```

```

PR 23-JUL-1999; 99DE-1034720.
XX (TIGR-) TIGR INST GENOMIC RES.
PA (QUIA-) QUIAGEN GMBH.
PA (GBFB) GES BIOLOGISCHES FORSCHUNG MBH.
PA (DKFZ-) DKFZ DEUT KREBSFORSCHUNGSZENTRUM.
PA (MEDI-) MEDIZINISCHE HOCHSCHULE HANNOVER.
XX
DR WPI; 2001-160596/17.
XX
PT New DNA encoding iron or heavy metal transporters, useful for promoting
PT growth of plants on polluted soil -
XX
PS Claim 3a; Page 7-8; 54pp; German.
XX
CC This invention describes novel DNA sequences (I) that encode iron or
CC heavy metal transporters (II). The invention also describes (1) a
CC recombinant expression vector (III) containing (I); (2) prokaryotic or
CC eukaryotic cells (III) transformed with (I) or (III); (3) production of
CC (II) by culturing (III); (4) (partial) expression products (IV) of (I)
CC and synthetic proteins or peptides with the same sequences; (5)
CC antibodies (Ab) specific for (IV); (6) hybridoma cells that produce
CC monoclonal Ab; and (7) transgenic plants that contain (III). The iron or
CC heavy metal transporters encoded by (I) transport heavy metal ions across
CC cell walls and regulate homeostasis of trace elements. (I), and their
CC fragments are useful for: (1) expression of (II); (2) as probes and
CC primers for detection, isolation and amplification of full length cDNA
CC sequences; and (3) producing transgenic plants. (III) are used to promote
CC growth, development and yield of plants, particularly Leguminosae,
CC especially when growing in soil polluted by heavy metals injurious to
CC plants. They also improve homeostasis of iron and trace elements. Host
CC cells that express the iron or heavy metal transporters take up heavy
CC metals, so reduce pollution of soil and release iron and other trace
CC metals which improves soil quality and protect plants against pollutants
CC from the soil.
XX
SQ Sequence 2586 BP; 563 A; 857 C; 742 G; 421 T; 3 other;
XX
Alignment Scores:
Pred. No.: 0.000111 Length: 2586
Score: 152.00 Matches: 68
Percent Similarity: 39.27% Conservative: 51
Best Local Similarity: 22.44% Mismatches: 142
Query Match: 7.43% Indels: 42
DB: Gaps: 12
US-09-889-746-2 (1-400) x AAF26411 (1-2586)
Qy 34 ValGluLeuGluProValValIleThrIleAspLysSerGlyMetAlaLeuAlaAsnArg 53
Db 388 ATCGCCCTGCAGCAGGTGACCGTCAGCGCCACCCGCGCAGC-----GCACAGGAC 435
Qy 54 IleThrGlnMetProHisThrThrLysValIleTyrGluGluGlnIleGlnGlnAla 73
Db 436 GTCAGCCAGGTGCCGACGACGTGTACGGTGCAGCCCTGAGCACTGACGCGGCGAGAAC 495
Qy 74 ThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSer 93
Db 496 GTGAACAATATCCAG-----GATCTGTGCGCTACGAGCGCGGGTTCGGTA--- 543
Qy 94 SerGlyThrThrSerAsnPheGly-----GlnThrMetHisGlyArg 107
Db 544 TCCGGAACCGGCCAACGCGCGGCTGAACCGCTACACATCCCGTGTATCGACGCGGAG 603
Qy 108 GlnValGlnPheLeuLeuAsnGlyValProLeuThrGlySer-----ArgAsp 123
Db 604 CGGATCTCCACCAAGTCAGCGCGTGTGATCCCGCAGCAGCTCTTCTACGGCCCTTAC 663
Qy 124 IleSerArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSer 143
Db 664 GCCCAGACCCGCGCAACTAGCTCGACCGGAAATCGTCAAGCGTGTGAAATCTCCGCG 723
Qy 144 GlyAlaThrSerIle---TyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThr 162

```





Qy 213 ArgThrThrGlyAlaPheAspAlaSerGlyLysArgIleAlaProGluProAlaGln 232  
 Db 604 CCGATCGGGTGTATTACGCCAGACATGCT-----GAAACCCGCGCG 648  
 Qy 233 ThrAspLysGlnAspSerLysSerLeuSerValAsnThrAsnValAspTyrGlnLeuAsp 252  
 Db 649 AATGACGAGTCCATTAAATACATGCTGGCGAAGGAGC-----TGGCAAAATGAT 699  
 Qy 253 AspLysGlnAsnIleAsnLeuAlaLeuThrHisTyrAsnAsp 266  
 Db 700 TCAGCCCATCTCTGCGCGGTTTAGTGCCTTACTACAACAC 741

RESULT 17  
 AAV81945  
 ID AAV81945 standard; DNA; 18831 BP.  
 XX AC AAV81945;  
 XX DT 19-OCT-1999 (first entry)  
 XX DE S. putrefaciens PKS-like cluster partial fragment.  
 XX KW Polyketide-like synthesis; PKS; PKS-like gene; PUFA; DHA; transgenic;  
 KW EPA; oil; dietary supplement; infant feeding formulation; malnutrition;  
 KW intravenous feeding formulation; cooking oil; fat; anti-inflammatory;  
 KW cholesterol; open reading frame; ORF; ss.  
 XX OS Shewanella putrefaciens.  
 XX FH Key Location/Qualifiers  
 FT CDS 6121..8106  
 FT /tag= a  
 FT /product= "ORF 2"  
 FT /note= "see AAW89396"  
 FT complement (8186..9016)  
 FT /tag= b  
 FT /product= "ORF 3"  
 FT /note= "see AAW89397"  
 FT 9681..12593  
 FT /tag= c  
 FT /product= "ORF 4"  
 FT /note= "see AAW89398"  
 FT 13040..13906  
 FT /tag= d  
 FT /product= "ORF 5"  
 FT /note= "see AAW89399"  
 FT 13906..18831  
 FT /tag= e  
 FT /product= "partial ORF 6"  
 FT /note= "see AAW89400; the complete sequence coding for  
 the ORF 6 is missing; ORF 6 is partially  
 encoded by the present sequence and sequence  
 shown in AAZ00331 (see AAY27276)"  
 FT misc\_feature 18831  
 FT /tag= f  
 FT /note= "nucleotides 18832 to 20019 (1188 base pairs)  
 of the PKS-like cluster sequence are missing  
 (given as a blank page in the specification);  
 the sequence continues from nucleotide 20020  
 and is provided in AAZ00331"

XX W09855625-A1.  
 XX PD 10-DEC-1998.  
 XX PF 04-JUN-1998; 98WO-US11639.  
 XX PR 04-JUN-1997; 97US-0048650.  
 XX PA (CALJ ) CALGENE LLC.

PI Pacciotti D, Lässner M, Metz JG;  
 XX WPI; 1999-070271/06.  
 DR N-PSDB; AAZ00331.  
 DR P-PSDB; AAW89396, AAW89397, AAW89398, AAW89399, AAW89400.  
 XX New nucleic acid encoding polyketide-like synthesis enzymes of  
 PT *Vibrio marinus* - and transformed plants and microbes that produce  
 FT polyunsaturated fatty acids, useful as pharmaceuticals and food  
 PT supplements  
 XX Claim 17; Fig 4; 153pp; English.  
 XX The invention provides polyketide-like synthesis (PKS)-like genes that  
 CC are used for the production of long chain polyunsaturated fatty acid  
 CC (PUFA) productions. Genes responsible for eicosapentenoic acid (EPA)  
 CC production in *Shewanella putrefaciens* and novel genes associated with  
 CC the production of docosahexanoic acid (DHA) in *Vibrio marinus* are used  
 CC to generate transgenic plants that can express transgenes encoding  
 CC PKS-like genes associated with PUFA production. The PKS-like genes are  
 CC used to transform plants and microbial cells to give recombinants having  
 CC altered contents of PUFA (specifically DHA and EPA). Oils from these  
 CC plants are useful as dietary supplements (in infant feeding  
 CC formulations, to give a PUFA profile closer to that of human milk; for  
 CC treating malnutrition; in intravenous feeding formulations; in cooking  
 CC oils, fat etc.), also as anti-inflammatory agents and for reducing  
 CC cholesterol levels. Fragments from the genes are useful as probes to  
 CC isolate related molecules or to detect organisms that express PKS-like  
 CC genes. The method facilitates large scale production of PUFA by providing  
 CC new pathways for their synthesis or suppressing interfering pathways.  
 CC Expression of PUFA in seeds allows simple recovery, as oil which can be  
 CC engineered to have a particular PUFA profile. Expression in microbes also  
 CC allows simple recovery and control of PUFA profile and is not subject to  
 CC external variables such as weather or food supply. The present sequence  
 CC represents a PKS-like cluster partial fragment from *S. putrefaciens*. The  
 CC sequence continues in AAZ00331 from nucleotide 20020. The sequence  
 CC representing nucleotides 18832 to 20020 (corresponding to 1188 basepairs  
 CC is not given in the specification with the information that it was  
 CC missing at the time of publication). The whole *S. putrefaciens* PKS-like  
 CC cluster comprises comprising open reading frames (ORFs) 3, 4, 5, 6, 7, 8  
 CC and 9. The present sequence encodes ORF 2, 3, 4, 5 and a partial  
 CC fragment of ORF 6.  
 XX SQ Sequence 18831 BP; 5760 A; 4229 C; 3822 G; 5020 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.00323 Length: 18831  
 Score: 149.00 Matches: 64  
 Percent Similarity: 46.03% Conservativity: 52  
 Best Local Similarity: 25.40% Mismatches: 83  
 Query Match: 7.28% Indels: 54  
 DB: 20 Gaps: 12

US-09-889-746-2 (1-400) x AAV81945 (1-18831)  
 Qy 3 HisSerHisTyrPheGlnTrpLeuSerLeuProLeu----- 14  
 Db 9660 TATACAATTATTATTAAGGGAATGATGATGTTTAAATTCAAAACCTTCGCGCTCAGTC 9719  
 Qy 15 -----LeuSerValAlaValThrGlnLeuTyrAlaGlnPro-----Asn 28  
 Db 9720 AAATCTGCCATATCCGCGGCTTAACAGCTCGCTA---GCTATGCTGTGTTTTCAGAA 9776  
 Qy 29 GluSerLeuProThrValGluLeuGluProValValIleThrIleAspLysSerGlyMet 48  
 Db 9777 GAAACTGCTGCTGAAGAACAAATAGAAAGATCGCAGTGACC-----GGATCG 9824  
 Qy 49 AlaLeuAlaAsnArgIleThrGlnMetProHisThrThrLysValIleTyrGluGln 68  
 Db 9825 CGAATCGCTAAAGCAGAGCTAACTCAACAGCTCCAGCTCGTCAGCCCTTTCAGCCGAA 9884  
 Qy 69 IleGlnGluGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIlePro 88

Db 9885 CTGACAAA-----TTTGGTATCAAGATTAGTAGGCTACTAGCAGATTA---CCT 9935  
 Qy 89 SerLeuGlyVal-----SerserGlyThr 96  
 Db 9936 GCATTGGGCAACCAACTATTATTGGTAATACATAGCACTCAAGCGAGGTGT 9995  
 Qy 97 ThrSerAspGlyValThrMetHsGlyArgGlnValGlnPheLeuLeuAngly--- 115  
 Db 9996 AGCTCAGACAGACTTGGTGTAGTGTACAGACACTTAGTATTAGTCAACGGTTAAG 10055  
 Qy 116 -----ValProLeuThrGlySerArgAspLieserArgGlnLeuLeuSerLieser 132  
 Db 10056 CGGTACGTGGCCGCCCAACCGGCTGAGTGAAGTA-----GATTGTCACTATACCA 10109  
 Qy 133 ProAsnGlnValAlaArgLiegValLeuSer---GlyAlaThrSerLiegGlySer 151  
 Db 10110 ACTAGCATGATCTCCGAGAGTGTGATGTACCGCGGCTCTTCAACATTTATGTTGG 10169  
 Qy 152 GlyAlaThrGlyGlyLeuLeuLeuValThrLysSerAspLeuGluGluGlnPhe 171  
 Db 10170 GACGCTGTATCAGGTGTATCAAGTATCTTAAGAAAGACTTGAAGGCTTGAAGTT 10229  
 Qy 172 GlnThrArgLiegGlyValHsGlySerLysLeuSerSerGluGlyLiegGlyThrGlnVal 191  
 Db 10230 AACGACGT-----ACTAGCGGTTCTTAAGTGAAGTGAAGTGAAGTGAAG 10274  
 Qy 192 -----GlyGlnSerValAlaGlyValSerGluAsnGlyVal 204  
 Db 10275 CACTCTTTGACATTTTGGTGTGCAAAAGCTTGCAGATGAGAGCTGTATGT- AACCTT 10333  
 Qy 205 LeuAlaArgLeuAspValAspValArgThrThrGly 216  
 Db 10334 CTACGCAAGTTATGAAAGTCAAAAGAAAGTATGGC 10369  
 RESULT 18  
 ID AAT34137 standard; DNA; 37895 BP.  
 AC AAT34137;  
 XX  
 XX AAT34137;  
 DT 30-JAN-1997 (first entry)  
 XX  
 DE Gene for biosynthetic enzyme group of icosapentaenoic acid synthase.  
 XX  
 KM Icosapentaenoic acid synthase; EPA; drugs; agrochemicals;  
 KM foodstuffs; animal feed; lipid balance correction; antihypertensive;  
 KM antinflammatory; anticancer agent; ss.  
 XX  
 OS Shewanella putrefaciens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 6121..8103  
 FT /tag= a  
 FT /product= Icosapentaenoic acid biosynthetic enzyme  
 FT complement (9016..8186)  
 FT /tag= b  
 FT /product= Icosapentaenoic acid biosynthetic enzyme  
 FT 9681..12590  
 FT /tag= c  
 FT /product= Icosapentaenoic acid biosynthetic enzyme  
 FT 13040..13903  
 FT /tag= d  
 FT /product= Icosapentaenoic acid biosynthetic enzyme  
 FT 13906..22173  
 FT /tag= e  
 FT /product= Icosapentaenoic acid biosynthetic enzyme  
 FT 22176..24515  
 FT /tag= f  
 FT /product= Icosapentaenoic acid biosynthetic enzyme  
 FT 24518..30529  
 FT /tag= g  
 FT /product= Icosapentaenoic acid biosynthetic enzyme  
 FT 30730..32358  
 CDS

FT /tag= h  
 FT /product= Icosapentaenoic acid biosynthetic enzyme  
 FT CDS 32753..34327  
 FT /tag= i  
 FT /product= Icosapentaenoic acid biosynthetic enzyme  
 PN W09621735-A1.  
 PD 18-JUL-1996.  
 PR 12-JAN-1996; 96WO-JP00030.  
 PR 13-JAN-1995; 95JP-0004299.  
 PA (SAGA ) SAGAMI CHEM RES CENTRE.  
 PI Kato S, Kondo K, Yamada A, Yazawa K;  
 DR WPI, 1996-342288/34.  
 DR P-PDB; AAR99458, AAR99459, AAR99460, AAR99461, AAR99462, AAR99463,  
 DR AAR99464, AAR99465, AAR99466.  
 XX  
 PT Production of icosapentaenoic acid using transformed E. coli - uses  
 PT DNA coding for icosapentaenoic acid synthase derived from Shewanella  
 strain  
 XX  
 PS Claim 1; Page 19-40; 145pp; English.  
 CC The DNA sequence encoding the biosynthetic enzymes of  
 CC Icosapentaenoic acid (EPA) can be used to transform Escherichia coli.  
 CC The DNA sequence allows efficient microbial production of EPA, which  
 CC is a raw material for drugs, agrochemicals, foods and animal  
 CC feedstuffs. EPA is also useful for lipid balance correction and as  
 CC an antihypertensive, antinflammatory and anticancer agent.  
 XX  
 SQ Sequence 37895 BP; 11266 A; 8624 C; 8032 G; 9972 T; 1 other;  
 Alignment Scores:  
 Pred. No.: 0.00835 Length: 37895  
 Score: 149.00 Matches: 64  
 Percent Similarity: 46.02% Conservatve: 52  
 Best Local Similarity: 25.40% Mismatches: 83  
 Query Match: 7.28% Indels: 54  
 DB: 17 Gaps: 12  
 US-09-889-746-2 (1-400) x AAT34137 (1-37895)  
 Qy 3 HsSerHisLeuThrPheGlnThrLeuSerLeuProLeu----- 14  
 Db 9660 TATTAACATTTATATTAAGGAATGATATGTTTAAATTCAAACTTCGCGCTCAGTC 9719  
 Qy 15 -----LeuSerValAlaValThrGlnGlnLeuTyAlaGlnPro-----Asn 28  
 Db 9720 AAACCTGCATATCCGACGCTTAACGCTCGCTA--GCTATGCTGTGTTTTCGAGAA 9776  
 Qy 29 GluSerLeuProThrValGluLeuGluProValAlaLeuThrLieserGlyMet 48  
 Db 9777 GAACCTGCTGCTGAGAAACAATAAGAAAGTCGCACTAC-----GGATCG 9824  
 Qy 49 AlaLeuAlaAsnArgLiegThrGlnMetProHsThrThrLysValLiegGluGluGln 68  
 Db 9825 CGAATCGCTAAAGCAGAGTCACTCAACGAGCTCCAGCTGAGCTTCAGCGAAGAA 9884  
 Qy 69 LiegGluGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaLeuLeuPro 88  
 Db 9885 CTGACAAA-----TTTGGTATCAAGATTAGTAGGCTACTAGCAGATTA---CCT 9935  
 Qy 89 SerLeuGlyVal-----SerserGlyThr 96  
 Db 9936 GCATTGGGCAACCAACTATTATTGGTAATACATAGCACTCAAGCGAGGTGT 9995  
 Qy 97 ThrSerAspGlyValThrMetHsGlyArgGlnValGlnPheLeuLeuAngly--- 115

```

Db 9996 AGCTCAGCAGACTTGGCTGCTAGGTGCTAACAGACCTTAGTATTAGTCAACGGTAAG 10055
Qy 116 -----ValProLeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAsn 132
Db 10056 CGCTACGTGGCGGCAACCGGCTCAGCTGAGGTA-----GATTGTGCAACTATACCA 10109
Qy 133 ProAsnGlnValAlaArgIleGluValLeuSer---GlyAlaThrSerIleTyrGlySer 151
Db 10110 ACTAGCATGATCTCGGAGTTCAGATTGPAACCGCGGCTTCAGCAATTTATGGTTGG 10169
Qy 152 GlyAlaThrGlyGlyLeuIleAsnIleValThrLysSerAspLeuGluGluGlnPhe 171
Db 10170 GACGCTGATCAGGTGTTATCAAGCTTATCCTTAAGAAGACTTTGAAGGCTTTGAGTT 10229
Qy 172 GluThrArgIleGlyValHisGlySerLysLeuSerSerGluGlyIleGlyTyrGlnVal 191
Db 10230 AACGCACGT-----ACTAGCGGTTCTACTGAAAGTGTAGGCACTCAAGAG 10274
Qy 192 -----GlyGlnSerValAlaGlyValSerGluAsnGlyAsnVal 204
Db 10275 CACTCTTTTGACATTTTGGGTGGTCAACAGTTGCAGATGCGAGTGGTAATGT-AACCTT 10333
Qy 205 LeuAlaArgLeuAspValAspTyrArgThrThrGly 216
Db 10334 CTACGAGGTTATGACGTACAAAGAGACTCATGGC 10369

RESULT 19
AAV00503
ID AAV00503 standard; DNA; 37895 BP.
XX AC AAV00503;
XX DT 03-JUL-1998 (first entry)
XX DE S. putrefaciens eicosapentaenoic acid biosynthesis gene cluster.
XX KW SCRC-2874; FERM BP-1625; eicosapentaenoic acid; EPA;
XX KW biosynthesis gene cluster; synthetase; ds.
XX OS Shewanella putrefaciens.
XX FH Location/Qualifiers
XX CDS 6121..8106
FT /tag= a
FT /note= "open reading frame 2"
FT complement (8183..9016)
FT /tag= b
FT /product= AAW37047
FT /note= "open reading frame 3"
FT 9681..12593
FT /tag= c
FT /product= AAW37048
FT /note= "open reading frame 4"
FT 13040..13906
FT /tag= d
FT /product= AAW37049
FT /note= "open reading frame 5"
FT 13906..22176
FT /tag= e
FT /product= AAW37050
FT /note= "open reading frame 6"
FT 22176..24518
FT /tag= f
FT /product= AAW37051
FT /note= "open reading frame 7"
FT 24518..30532
FT /tag= g
FT /product= AAW37052
FT /note= "open reading frame 8"
FT 30730..32361
FT /tag= h
FT /product= AAW37053
FT /note= "open reading frame 9"

```

```

FT CDS 32753..34330
FT /*tag= i
FT /note= "open reading frame 10"
XX WO9801565-A1.
XX 15-JAN-1998.
XX 09-JUL-1997; 97WO-JP02371.
XX 10-JUL-1996; 96JP-0180845.
XX (SAGA ) SAGAMI CHEM RES CENTRE.
XX Kato S, Kondo K, Yamada A, Yazawa K;
XX WPI; 1998-101060/09.
XX P-PSDB; AAW37047, AAW37048, AAW37049, AAW37050, AAW37051, AAW37052,
XX AAW37053.
XX Eicosapentaenoic acid produced by culture of transformed Escherichia
XX coli - containing an eicosapentaenoic acid synthetase gene derived
XX from the marine microorganism Shewanella
XX Claim 1; Pages 18-41; 110pp; Japanese.
XX
XX The present sequence is the Shewanella putrefaciens SCRC-2874
XX (FERM BP-1625) eicosapentaenoic acid (EPA) biosynthesis gene
XX cluster.
XX A novel EPA (useful in drugs, pesticides, foods and feedstuffs) is
XX encoded by synthetase enzyme gene sequences comprising parts of the
XX full sequence of the synthetase gene from the marine microorganism
XX S. putrefaciens SCRC-2874 (FERM BP-1625), in which at least 1 of
XX the 9 open reading frames (ORF) (numbered 2-10) in the gene have
XX been deleted. In particular the gene sequences comprising the
XX following parts of the full gene:
XX (1) bases 8081-9441, 12314-13084 and 13889-32520;
XX (2) bases 8081-9441, 12314-13084, 13889-32520 and 34627-35559;
XX (3) bases 8081-9441, 12314-13084 and 13889-35559;
XX (4) bases 8081-9441, 9681-13084 and 13889-32520;
XX (5) bases 8081-9441, 9681-13084, 13889-32520 and 34627-35564; and
XX (6) bases 8081-9441, 9681-13084 and 13889-35564, are claimed.
XX
SQ Sequence 37895 BP; 11266 A; 8623 C; 8032 G; 9972 T; 2 other;

Alignment Scores:
Pred. No.: 0.00835 Length: 37895
Score: 149.00 Matches: 64
Percent Similarity: 46.03% Conservative: 52
Best Local Similarity: 25.40% Mismatches: 83
Query Match: 7.28% Indels: 54
DB: 19 Gaps: 12

US-09-889-746-2 (1-400) x AAV00503 (1-37895)
Qy 3 HisSerHisTyrPheGlnTrpLeuSerLeuProLeu----- 14
Db 9660 TATAACAATTATATTAGGAATGATGATGTTTAAATTCAAAACCTTTCGGCTCAGTC 9719
Qy 15 -----LeuSerValAlaValThrGlnLeuTyrAlaGlnPro-----Asn 28
Db 9720 AAACCTTGCCATATCGCAGGCTTAACAGCTCGCTA---GCTATGCCCTGTTTTCGAGAA 9776
Qy 29 GluSerLeuProThrValGluLeuGluProValIleThrIleAspLysSerGlyMet 48
Db 9777 GAAACTGCTGCTGAAGAACAAATAGAAAGAGTCGACAGTGACC-----CGATCG 9824
Qy 49 AlaLeuAlaAsnArgIleThrGlnMetProHisThrThrLysValIleTyrGluGln 68
Db 9825 CGAATCGCTAAAGCAGAGACTAACTCAACAGCTCGTCAGCTTTTCAGCCCAAGAA 9884
Qy 69 IleGlnGluGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIlePro 88

```

Db 9885 CTGACAAA-----TTTGGTAATCAAGATTAGTAGCTACTAGCAGAAATTA---CCT 9935  
 QY 89 SerLeuGlyVal-----SerSerGlyThr 96  
 Db 9936 GCGATTGGTGCAACCAACTATTATTGTATATACATAGCAACTCAACGCGAGGTGT 9995  
 QY 97 ThrSerAspPheGlyGlnThrMetHisGlyArgGlnValGlnPheLeuLeuAngly--- 115  
 Db 9996 AGCTCAGCAGACTGGCTGCTCTAGGTGTCAACGAACTTAGATTAGTCAACGCGTAAG 10055  
 QY 116 -----ValProLeuThrGlySerArgAspHisSerArgGlnLeuLeuSerIleAsn 132  
 Db 10056 CGGTACGTCGCCGCCCAACCGGCTCAGCTAGAGTA-----GATTGTCAACTATACCA 10109  
 QY 133 ProAsnGlnValAlaArgIleGlnValLeuSer---GlyAlaThrSerIleTyrglySer 151  
 Db 10110 ACTAGCATATATCGCGAGTGTGAGATTGTACCGCGGCTGCTCAGCAATTTATGTTGG 10169  
 QY 152 GlyAlaThrGlyGlyLeuIleAsnIleValThrLySerAspLeuGlnGlnPhe 171  
 Db 10170 GACGCTGTATCAAGTGTATCAAGTATCTTAAAGAACCTTTGACGCTTGAAGTTT 10229  
 QY 172 GluThrArgIleGlyValHisGlySerLyLeuSerSerGlnGlyIleGlyTyrglnVal 191  
 Db 10230 AACGACGTT-----ACTAGCGGTTCTACTAAAGTGTAGGCGCTCAAGAG 10274  
 QY 192 -----GlyGlnSerValAlaGlyValSerGlnAsnGlyAsnVal 204  
 Db 10275 CACTCTTTGACATTTTGGTGTGCAAAAGCTTGACAGATGAGAGTGTATATG-MACTT 10333  
 QY 205 LeuAlaArgLeuAspValAspTyrgTrpThrGly 216  
 Db 10334 CTACGCGATTATGACGTACAAAGAAAGATCATGCG 10369  
 RESULT 20  
 AAA71518  
 ID AAA71518 standard; DNA; 37895 BP.  
 AC AAA71518;  
 XX 11-DEC-2000 (first entry)  
 DT  
 XX Shewanella putrefaciens PKS gene cluster encoding ORF3 to ORF9.  
 DB  
 XX PKS pathway; polyunsaturated long chain fatty acid; plant; transgenic;  
 KM polyketide-like synthesis; PUFA; dietary supplement; intravenous feeding;  
 KM malnutrition; cooking oil; cooking fat; margarine;  
 KM docosahexenoic acid production; eicosapentenoic acid production; ds.  
 XX  
 XX Shewanella putrefaciens.  
 OS  
 PN WO200042195-A2.  
 XX  
 PD 20-JUL-2000.  
 PF 14-JAN-2000; 2000WO-US00956.  
 XX  
 PR 14-JAN-1999; 99US-0231899.  
 XX  
 PA (CALJ ) CALGENE LLC.  
 PI Faciolletti D, Metz JG, Lassner M;  
 XX  
 DR WPI; 2000-476063/41.  
 PT New DNA sequences encoding for polyketide (PK)-like synthesis pathway  
 PT genes from Shewanella, Vibrio and Schizochitrium, useful for creating  
 PT transgenic plants that express poly-unsaturated long chain fatty acids  
 PT  
 XX  
 PS Example 1; Fig 4A; 302pp; English.  
 CC This invention describes novel DNA sequences encoding for polyketide

CC (PK)-like synthesis (PKS-like) pathway genes from Shewanella, Vibrio  
 CC and Schizochitrium. The nucleic acids are useful for isolating related  
 CC molecules or in methods to detect organisms expressing the PKS-like  
 CC genes. They are also useful for creating transgenic plants that express  
 CC poly-unsaturated long chain fatty acids. The poly-unsaturated long chain  
 CC fatty acids produced recombinantly are useful as dietary supplements for  
 CC patients undergoing intravenous feeding or for preventing or treating  
 CC malnutrition. The poly-unsaturated long chain fatty acids can also be  
 CC incorporated into cooking oils, fats or margarine formulated so that in  
 CC normal use the recipient receives a desired amount of poly-unsaturated  
 CC long chain fatty acids. The nucleic acids are also useful in large scale  
 CC production of docosahexenoic acid and eicosapentenoic acid, and for the  
 CC modification of the fatty acid profile of host cells and edible plant  
 CC tissues and/or plant parts. Transgenic production of polyunsaturated  
 CC fatty acids in particular host cells allows quicker purification from  
 CC natural sources such as fish or plants. This sequence represents the  
 CC Shewanella putrefaciens DNA sequence encoding the PKS gene cluster  
 CC incorporating ORF3 to ORF9.  
 XX  
 SQ Sequence 37895 BP; 11266 A; 8624 C; 8031 G; 9972 T; 2 other;  
 Alignment Scores:  
 Pred. No.: 0.00835 Length: 37895  
 Score: 149.00 Matches: 64  
 Percent Similarity: 46.03% Conservative: 52  
 Best Local Similarity: 25.40% Identities: 83  
 Query Match: 7.28% Mismatches: 54  
 DB: 21 Gaps: 12  
 US-09-889-746-2 (1-400) x AAA71518 (1-37895)  
 QY 3 HisSerHisTyPheGlnThrProLeuSerLeuProLeu----- 14  
 Db 9660 TATPACATATATATTAAGGAATGAGTATGTTTAAATCAAACTTCGGCTCAGTC 9719  
 QY 15 -----LeuSerValAlaValThrGlnGlnLeuTyrglnPro-----Asn 28  
 Db 9720 AAACCTGCATATCCGACGAGCTTAACAGCCCTGCTA---GCTATGCTGTGTTTGCAGAA 9776  
 QY 29 GluSerLeuProThrValGlnLeuGluProValIleThrIleAspIleSerGlyMet 48  
 Db 9777 GAAACGCTGCTGAAGAACAAATAGAAAGTGCAGTAC-----GATTCG 9824  
 QY 49 AlaLeuAlaAsnArgIleThrGlnMetProHisThrThrLyValIleTyrglnGln 68  
 Db 9825 CGAATGCTTAAAGCAGAGTACTCAACGCTCCAGTGTACGCTTTCAGCCGAAGAA 9884  
 QY 69 IleGlnGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIlePro 88  
 Db 9885 CTGACAAA-----TTTGGTAATCAAGATTAGTAGCTACTAGCAGAAATTA---CCT 9935  
 QY 89 SerLeuGlyVal-----SerSerGlyThr 96  
 Db 9936 GCGATTGGTGCAACCAACTATTATTGTATATACATAGCAACTCAACGCGAGGTGT 9995  
 QY 97 ThrSerAspPheGlyGlnThrMetHisGlyArgGlnValGlnPheLeuLeuAngly--- 115  
 Db 9996 AGCTCAGCAGACTGGCTGCTCTAGGTGTCAACGAACTTAGATTAGTCAACGCGTAAG 10055  
 QY 116 -----ValProLeuThrGlySerArgAspHisSerArgGlnLeuLeuSerIleAsn 132  
 Db 10056 CGGTACGTCGCCGCCCAACCGGCTCAGCTAGAGTA-----GATTGTCAACTATACCA 10109  
 QY 133 ProAsnGlnValAlaArgIleGlnValLeuSer---GlyAlaThrSerIleTyrglySer 151  
 Db 10110 ACTAGCATATATCGCGAGTGTGAGATTGTACCGCGGCTGCTCAGCAATTTATGTTGG 10169  
 QY 152 GlyAlaThrGlyGlyLeuIleAsnIleValThrLySerAspLeuGlnGlnPhe 171  
 Db 10170 GACGCTGTATCAAGTGTATCAAGTATCTTAAAGAACCTTTGACGCTTGAAGTTT 10229  
 QY 172 GluThrArgIleGlyValHisGlySerLyLeuSerSerGlnGlyIleGlyTyrglnVal 191

Db 10230 AACGCACGT-----ACTACGGTCTTACTGAAAGTGTAGGCTCAAGAG 10274  
Qy 192 -----GlyGlnSerValAlaGlyValSerGluAenGlyAenVal 204  
Db 10275 CACTCTTTTGACATTTTGGGTGGTCAACAGTTGCAGATGGAGCTGTAATGT-AACCTT 10333  
Qy 205 LeuAlaArgLeuAspValAspTyrArgThrThrGly 216  
Db 10334 CTACGCAGGTTATGAACGTACAAAAGAGCTATGGC 10369  
RESULT 21  
AAH47833  
ID AAH47833 standard; DNA; 37895 BP.  
XX  
XX AAH47833;  
AC  
XX  
DT 22-OCT-2001 (first entry)  
XX  
DE S. putrefaciens eicosapentanoic acid synthesis gene cluster.  
XX  
KW Cyanobacterium; eicosapentanoic acid; EPA; plasmid; ds.  
XX  
OS Shewanella putrefaciens.  
XX  
PN JP2001145490-A.  
XX  
PD 29-MAY-2001.  
XX  
PF 19-NOV-1999; 99JP-0329169.  
XX  
PR 19-NOV-1999; 99JP-0329169.  
XX  
XX (SAGA ) SAGAMI CHEM RES CENT.  
PA (BIOI-) BIOINDUSTRY KYOKAI SH.  
PA (KEIZ-) KEIZAI SANGYOSHO SANGYO GIJUTSU SOGO KEN.  
XX  
XX WPI; 2001-406151/43.  
DR  
XX  
PT A plasmid in which eicosapentanoic acid biosynthesis gene group is  
PT cloned and used to transform cyanobacterium so that it produces  
PT eicosapentanoic acid  
XX  
PS Disclosure; Page 6-18; 62pp; Japanese.  
XX  
CC The invention relates to a plasmid prepared by cloning the Shewanella  
CC putrefaciens SCRC-2874 (FERM BP-1625) eicosapentanoic acid (EPA)  
CC synthesis gene cluster into a broad host vector. The plasmid is used  
CC to transform cyanobacterium and produce EPA.  
XX  
SQ Sequence 37895 BP; 11266 A; 8625 C; 8032 G; 9971 T; 1 other;  
  
Alignment Scores:  
Pred. No.: 0.00835 Length: 37895  
Score: 149.00 Matches: 64  
Percent Similarity: 46.03% Conservative: 52  
Best Local Similarity: 25.40% Mismatches: 83  
Query Match: 7.28% Indels: 54  
DB: 12 Gaps: 12  
  
US-09-889-746-2 (1-400) x AAH47833 (1-37895)  
Qy 3 HisSerHisTyrPheGlnTrpLeuSerLeuProLeu----- 14  
Db 9660 TATAACAATTATTAAGGAATGAGTATGTTTAAATTCAAAACCTTTTCGGCTCAGTC 9719  
Qy 15 -----LeuSerValAlaValThrGlnLeuTyrAlaGlnPro-----Aen 28  
Db 9720 AAACCTGGCATATCCGACGGCTTAACAGCTCGCTA---GCTATGCTGTTTTCGAGAA 9776  
Qy 29 GluSerLeuProThrValGluLeuGluProValValIleThrIleAspLysSerGlyMet 48  
Db 9777 GAAACTGCTGCTGAAGAACAAATAGAAAGAGTCGAGTGACC-----GGATCG 9824

Qy 49 AlaLeuAlaAenArgIleThrGlnMetProHisThrThrLysValIleTyrGluGln 68  
Db 9825 CGAATCGCTAAAGCAGAGCTAACTCAACACGCTCCAGCTCGTCAGCCCTTTCAGCCGAAGA 9884  
Qy 69 IleGlnGluGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIlePro 88  
Db 9885 CTGACAAA-----TTTGGTAAATCAGATTTAGGTAGCTACTAGCAGAAATTA---CCT 9935  
Qy 89 SerLeuGlyVal-----SerSerGlyThr 96  
Db 9936 GCTATTGGTCAACCAACACACTATTATTGGTAATAACAATACCACTCAAGCGCAGGTCTT 9995  
Qy 97 ThrSerAenPheGlyGlnThrMetHisGlyArgGlnValGlnPheLeuLeuAenGly--- 115  
Db 9996 AGCTCAGCAGACTTTCGCTGCTAGGTGCTAACACAGAACCTTAGTATTAGTCAACGGTAAG 10055  
Qy 116 -----ValProLeuThrGlySerArgAspIleSerArgGlnLeuAenSerIleAen 132  
Db 10056 CGTACGTTCGGCCCAACCGGCTCAGCTGAGTA-----GATTTGTCACTATATCA 10109  
Qy 133 ProAenGlnValAlaArgIleGluValLeuSer---GlyAlaThrSerIleTyrGlySer 151  
Db 10110 ACTAGCATGATCTCGGAGTTGAGATTGCTAAACCGCGGTGCTTCAGCAATTTATGGTTCCG 10169  
Qy 152 GlyAlaThrGlyGlyLeuIleAenIleValThrLysSerAspLeuGluGluGlnPhe 171  
Db 10170 GACGCTGTATCAGGTGTTATCAACGTTATCTCTTAAGAAGACTTTGAAGGCTTTGAGTTT 10229  
Qy 172 GluThrArgIleGlyValHisGlySerLysLeuSerSerGluGlyIleGlyTyrGlnVal 191  
Db 10230 AACGCACGT-----ACTAGCGTTCCTACTGAAGTGTAGGCTCAAGAG 10274  
Qy 192 -----GlyGlnSerValAlaGlyValSerGluAenGlyAenVal 204  
Db 10275 CACTCTTTTGACATTTTGGGTGGTCAACAGTTGCAGATGGACGCTGTAATGT-AACCTT 10333  
Qy 205 LeuAlaArgLeuAspValAspTyrArgThrThrGly 216  
Db 10334 CTACGCAGGTTATGAACGTACAAAAGAGTCAATGTC 10369  
RESULT 22  
AAQ51128  
ID AAQ51128 standard; DNA; 37913 BP.  
XX  
XX AAQ51128;  
AC  
DT 27-MAY-1994 (first entry)  
XX  
XX Eicosapentanoic acid (EPA) synthesis operon.  
XX  
XX EPA; eicosapentanoic acid synthetase; drug; anticoagulant;  
KW hypolipemic; hypoglycemic; antihypertensive; anticancer; pesticide;  
KW foodstuff; additive; ss.  
XX  
OS Shewanella putrefaciens.  
XX  
FH Key Location/Qualifiers  
FT CDS 6121..8103  
FT /\*tag= a  
FT /product= Enzyme used in EPA biosynthetic pathway.  
FT 9681..12590  
FT /\*tag= b  
FT /product= Enzyme used in EPA biosynthetic pathway.  
FT CDS 13905..18806  
FT /\*tag= c  
FT /product= Enzyme used in EPA biosynthetic pathway.  
FT CDS 18997..22170  
FT /\*tag= d  
FT /product= Enzyme used in EPA biosynthetic pathway.  
FT CDS 22173..24512  
FT /\*tag= e  
FT /product= Enzyme used in EPA biosynthetic pathway.  
FT CDS 24515..30151

```

FT      /tag= f
FT      /product= Enzyme used in EPA biosynthetic pathway.
FT      CDS      30727..32355
FT      /tag= g
FT      /product= Enzyme used in EPA biosynthetic pathway.
FT      CDS      32750..34345
FT      /tag= h
FT      /product= Enzyme used in EPA biosynthetic pathway.
PN      WO9323545-A.
XX
XX      25-NOV-1993.
XX
XX      14-MAY-1993; 93WO-JP00641.
XX
XX      15-MAY-1992; 92JP-0147945.
XX
XX      (SAGA ) SAGAMI CHEM RES CENTRE.
XX
XX      Kato S, Kondo K, Yamada A, Yazawa K,
XX      WPI: 1993-386577/48.
XX      P-PSDB; AAR42450-57.
XX
XX      Gene coding for eicoso-penta:enoic acid synthetase - is isolated
XX      from Pseudomonas, Alteromonas or Shewanella and used for
XX      recombinant prodn. of eicoso-penta:enoic acid
XX
XX      Example 3; Page 16-39; 106pp; Japanese.
XX
XX      EPA is useful as a drug, having anticoagulant, hypolipemic,
XX      hypoglycemic, antihypertensive and anticancer activity. It is also
XX      a pesticide and is useful as a nutritional feedstuff and animal feed
XX      additive.
XX
XX      Sequence 37913 BP, 11275 A, 8689 C, 7966 G, 9980 T, 3 other;
SQ
Alignment Scores:
Pred. No.: 0.00835 Length: 37913
Score: 149.00 Matches: 64
Percent Similarity: 46.03% Conservative: 52
Best Local Similarity: 25.40% Mismatches: 83
Query Match: 7.28% Indels: 54
DB: Gaps: 12
US-09-889-746-2 (1-400) x AAO51128 (1-37913)
QY      3 HisSerHisIyrPheGlnTrpLeuSerLeuProLeu----- 14
DB      9660 TATTAACAATTATTAAGGAGATGATGTTTAAATTCAAACTTTCGCGTCAGTC 9719
QY      15 -----LeuSerValAlaValThrGlnLeuTrpAlaGlnPro-----Asn 28
DB      9720 AAATCGCCATATCCGACGGCTTAACGCTCGCTA---GCTATGCTGTTTTCAGAA 9776
QY      29 GluSerLeuProThrValGluLeuGluProValValIleThrIleAspIleSerGlyMet 48
DB      9777 GAAACGCTGCTGAAGAACAAATAGAAAGTGGCAGTGACC-----GATTCG 9824
QY      49 AlaLeuAlaAsnArgIleThrGlnMetProHisThrThrValIleIleTrpGlnGln 68
DB      9825 CGAATCGCTAAACAGAGCTAACTCAACAGCTCCAGTCCGACCTTCAGCGAAGAA 9884
QY      69 IleGlnGlnIleAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIlePro 88
DB      9885 CTACACAAA-----TTTGGTAAATCAGATTGAGTAGCGTACTAGCAGAAATTA---CCT 9935
QY      89 SerLeuGlyVal-----SerSerGlyThr 96
DB      9936 GCATTGGTGCACCAACACTATTATTGTAATACAAATAGCAACTCAACGCGAGGTGTT 9995
QY      97 ThrSerAsnPheGlyGlnThrMetHisGlyArgGlnValGlnPheLeuLeuAsnGly--- 115

```

```

DB      9996 AGCTACAGACACTTGGCTGCTAGGTGCTAACAGAACCTTAGTATTACTCAACGGTAG 10055
QY      116 -----ValProLeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAsn 132
DB      10056 CGCTACGTGGCCGGCCACCGGCTCAGTGAAGTA-----GATTGGCAACTATACCA 10109
QY      133 ProAsnGlnValAlaArgIleGluValLeuSer---GlyAlaThrSerIleTrpGlySer 151
DB      10110 ACTAGCATGATCTCGCGAGTGAAGATTGTAACCGGGGTGCTTACGAAATTATAGGTTG 10169
QY      152 GlyAlaThrGlyGlyLeuIleAsnIleValThrIleSerAspLeuGlnGluGlnPhe 171
DB      10170 GACGCTGATACAGGTGTATCAACCTTATCAAGAACTTGAAGCTTGAAGT 10229
QY      172 GluThrArgIleGlyValHisGlySerIleSerSerGlyGlyIleGlyTrpGlnVal 191
DB      10230 AACGACGT-----ACTACGGCTTCTACTGAAAGTATGACCTCAAGAG 10274
QY      192 -----GlyGlnSerValAlaGlyValSerGluAsnGlnAsnVal 204
DB      10275 CACTCTTTGACATTTTGGTGGTGCAGAACGTTGCAGATGACGTGTATATGT-ACCCTT 10333
QY      205 LeuAlaArgLeuAspValAspTrpArgThrThrGly 216
DB      10334 CTACGCGAGTTATGACGTACAAAGAAAGTCAATGCG 10369
RESULT 23
AAS46247/C
ID      AAS46247 standard; DNA, 10731 BP.
XX
XX      AAS46247;
XX
XX      18-DEC-2001 (first entry)
XX
XX      DNA encoding novel mar regulated protein (NIMR) #16.
XX
XX      mar regulated polypeptide; NIMR; microbial infection; antibacterial; ds.
XX
XX      Escherichia coli.
XX
XX      WO200170776-A2.
XX
XX      27-SEP-2001.
XX
XX      08-MAR-2001; 2001WO-US07478.
XX
XX      10-MAR-2000; 2000US-188362P.
XX
XX      (TUFT ) TUFTS COLLEGE.
XX
XX      Levy SB, Barbosa TM, Alekshun MN;
XX
XX      WPI; 2001-602769/68.
XX
XX      P-PSDB; AAU29347.
XX
XX      Identifying compounds that modulate a newly identified mar regulated
XX      polypeptide activity, useful as antimicrobial compounds, involves
XX      contacting the polypeptide with a test compound -
XX
XX      Disclosure; Page 259-265; 526pp; English.
XX
XX      The invention relates to a method of identifying compounds that modulate
XX      a newly identified mar regulated (NIMR) polypeptide activity. The method
XX      comprises contacting an NIMR polypeptide with a test compound under
XX      interaction conditions, determining the ability of the compound to
XX      modulate the activity or expression of the polypeptide, and selecting the
XX      CC modulators. NIMR nucleic acids and polypeptides are used in the treatment
XX      of microbial infections, and in screening for modulators of NIMR
XX      expression and activity. These modulators can be used to reduce the
XX      infectivity of a microbe on a surface, and the virulence of a microbe in
XX      a subject suffering from an infection. AAS46232-AAS46278 represent
XX      Escherichia coli NIMR coding sequences of the invention.

```

Db	8300	-----GCGACACCCAGAAATACCAACTCC---	:::
Qy	294	LysGlyLeuSerLeuSerGluGlnProLysThrLysSerThrPheAsnIleAsnTyr	:::
Db	8276	GATTCTATACCCTCGAATATGGCATGAACCAACCGTCGTATCGGCAGACTAC	:::
Qy	314	-----HisHseAspLeuTrpGlyAenThrIleAsnThrAsnAlaTyrTy-ArgArg	:::
Db	8216	CGCGTGACCTGGAAACGGTGGTGGGATAACGGCGTGACCAACGCACTGGGTGCAGTAC	:::
Qy	332	GluLysGlyArg 335	
Db	8156	GAACACACCCGT 8145	
RESULT	24		
AAS82105			
ID	AAS82105	standard; cDNA; 2157 BP.	
XX	AAS82105;		
XX	13-FEB-2002	(first entry)	
DE	DNA encoding novel human diagnostic protein #17909.		
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
KW	food supplement; medical imaging; diagnostic; genetic disorder; es.		
OS	Homo sapiens.		
FN	WO200175067-A2.		
PD	11-OCT-2001.		
PF	30-MAR-2001; 2001WO-US08631.		
PR	31-MAR-2000; 2000US-0540217.		
PR	23-AUG-2000; 2000US-0649167.		
PA	(HYSE-) HYSEQ INC.		
Pf	Dmanac RT, Liu C, Tang YT;		
DR	WPI; 2001-639362/73.		
DR	P-PSDB; ABGI7918.		
PT	New isolated polynucleotide and encoded polypeptides, useful in		
PT	diagnostics, forensics, gene mapping, identification of mutations		
PT	responsible for genetic disorders or other traits and to assess		
PT	biodiversity -		
Claim	1; SEQ ID No 17909; 103pp; English.		
The	invention relates to isolated polynucleotide (I) and		
polypeptide	(II) sequences. (I) is useful as hybridisation probes,		
polymerase	chain reaction (PCR) primers, oligomers, and for chromosome		
and	gene mapping, and in recombinant production of (II). The		
polynucleotides	are also used in diagnostics as expressed sequence tags		
for	identifying expressed genes. (I) is useful in gene therapy techniq		
to	restore normal activity of (II) or to treat disease states involvin		
(II).	(II) is useful for generating antibodies against it, detecting o		
quantitating	a polypeptide in tissue, as molecular weight markers and		
a	food supplement. (II) and its binding partners are useful in medic		
imaging	of sites expressing (II). (I) and (II) are useful for treating		
disorders	Involving aberrant protein expression or biological activity		
The	polypeptide and polynucleotide sequences have applications in		
diagnostics,	forensics, gene mapping, identification of mutations		
responsible	for genetic disorders or other traits to assess biodiver		
and	to produce other types of data and products dependent on DNA and		
amino acid	sequences. AAS64197-AAS94564 represent novel human		
diagnostic	coding sequences of the invention.		
Note:	The sequence data for this patent did not appear in the printed		
specification,	but was obtained in electronic format directly from WITN		





CC fatty acids produced recombinantly are useful as dietary supplements for  
 CC patients undergoing intravenous feeding or for preventing or treating  
 CC malnutrition. The poly-unsaturated long chain fatty acids can also be  
 CC incorporated into cooking oils, fats or margarine formulated so that in  
 CC normal use the recipient receives a desired amount of poly-unsaturated  
 CC long chain fatty acids. The nucleic acids are also useful in large scale  
 CC production of docosahexaenoic acid and eicosapentaenoic acid, and for the  
 CC modification of the fatty acid profile of host cells and edible plant  
 CC tissues and/or plant parts. Transgenic production of polyunsaturated  
 CC fatty acids in particular host cells allows quicker purification from  
 CC natural sources such as fish or plants. This sequence represents the  
 CC Shewanella putrefaciens DNA sequence encoding the PKS gene cluster  
 CC ORF4 which is described in the method of the invention.  
 XX

SQ Sequence 2910 BP; 824 A; 610 C; 651 G; 825 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.000356 Length: 2910  
 Score: 147.50 Matches: 58  
 Percent Similarity: 47.79% Conservative: 50  
 Best Local Similarity: 25.66% Mismatches: 78  
 Query Match: 7.21% Indels: 41  
 DB: 21 Gaps: 10

US-09-889-746-2 (1-400) x AA71576 (1-2910)

QY 15 LeuSerValAlaValThrGlnGlnLeuYrAlaGlnProAsnGluSerLeuProThrVal 34  
 DB 61 TTAACAGCTCGCTAGCTAGCTGCTGTTTTCAGAA-----GAACTGCTCTGAAGAA 114  
 QY 35 GluLeuGluProValValIleThrLeuAspLysSerGlyMetAlaLeuAlaAsnArgIle 54  
 DB 115 CAAATAGAAAGATGCGCAGTGACC-----GGATCGCAATCGCTAAAGCAGAG 162  
 QY 55 ThrGlnMetProHisThrThrLysValIleTyrGluGluGlnIleGlnGlnAlaThr 74  
 DB 163 CTAACCTCAACAGCTCCAGCTGCTGAGCTTTTCAGCGGAAGAACTGACAAA-----TTT 216  
 QY 75 GlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyVal----- 92  
 DB 217 GGTAAATCAAGATTAGGTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCT 273  
 QY 93 -----SerSerGlyThrThrSerAsnPheGlyGln 102  
 DB 274 ACTATTATTGGTAATAACAACTAGCACTCAAGCGGAGGTGTTAGCTCAGCAGCTTGGCT 333  
 QY 103 ThrMetHisGlyArgGlnValGlnPheLeuLeuAsnGly-----ValProLeu 118  
 DB 334 CQTCTAGGTGCTAAACAGAACCTTAGTATTAGTCAACGGTAAGCGGTACGTTGCCGCCAA 393  
 QY 119 ThrGlySerArgAspLysSerArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArg 138  
 DB 394 CCGGCTCAGCTGAGTA-----GATTGTCACTATACCACTAGCATGATCTCGCA 447  
 QY 139 IleGluValLeuSer---GlyAlaThrSerIleTyrGlySerGlyAlaThrGlyGlyLeu 157  
 DB 448 GTTGAGATTGAACCGCGGTGCTTTCAGCAATTTATGTTTCGACGCTGTATCAGGTGTT 507  
 QY 158 IleAsnIleValThrLysSerAspLeuGluGlnPheGluThrArgIleGlyVal 177  
 DB 508 ATCAAGTTTATCTTAAAGAAGACTTTGAGGCTTTGAGTTTAAAGCCACGT----- 558  
 QY 178 HisGlySerLysLeuSerGluGlyIleGlyTyrGlnVal----- 191  
 DB 559 -----ACTAGCGGTCTACTGAAAGTGTAGGCACTCAAGAGCACTCTTTTCACATTTTG 612  
 QY 192 ---GlyGlnSerValAlaGlyValSerGluAsnGlyAsnValLeuAlaArgLeuAspVal 210  
 DB 613 GGTGGTGCACAAAGCTTTCAGATGACGCTGTTGTAATGT-AACTTCTACGAGGTTATGAACG 671  
 QY 211 AspTyrArgThrThrGly 216  
 DB 672 TACAAAGAAGATCATGGC 689

RESULT 26

AAZ54328  
 ID AAZ54328 standard; DNA; 2127 BP.

XX AC AAZ54328;

XX DT 21-MAR-2000 (first entry)

XX DE Neisseria meningitidis ORF 760 partial DNA sequence SEQ ID NO:2605.

XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;  
 KW antibacterial; gene therapy; ds.

XX OS Neisseria meningitidis.

XX PN WO957280-A2.

XX PD 11-NOV-1999.

XX PF 30-APR-1999; 99WO-US09346.

XX PR 01-MAY-1998; 98US-0083758.

XX PR 31-JUL-1998; 98US-0094889.

XX PR 02-SEP-1998; 98US-0098994.

XX PR 02-SEP-1998; 98US-0099062.

XX PR 09-OCT-1998; 98US-0103749.

XX PR 09-OCT-1998; 98US-0103794.

XX PR 09-OCT-1998; 98US-0103796.

XX PR 25-FEB-1999; 99US-0121528.

XX PA (CHIR ) CHIRON CORP.

XX PA (GENO-) INST GENOMIC RES.

XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;

XX PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;

XX PI Tettelin H, Venter JC;

XX DR WPI; 2000-062150/05.

XX DR P-PSDB; AAY75566.

XX PT Novel Neisserial polypeptides predicted to be useful antigens for

XX PT vaccines and diagnostics -

XX PS Claim 7; Page 1235; 1453pp; English.

XX CC AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941

XX CC represent novel Neisseria meningitidis and N. gonorrhoeae polynucleotides

XX CC and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent

XX CC PCR primers used in the exemplification of the present invention. The

XX CC polypeptides, the polynucleotides, antibodies and compositions of

XX CC the invention can be used as vaccines, as diagnostic reagents, and as

XX CC immunogenic compositions. The polypeptides can be used in the

XX CC manufacture of medicaments for treating or preventing infection due to

XX CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the

XX CC presence of Neisseria bacteria, or to raise antibodies. They may also

XX CC be used to screen for agonists or antagonists, which may themselves

XX CC have use as antibacterial agents. The polynucleotides of the invention

XX CC may also be used in gene therapy protocols.

SQ Sequence 2127 BP; 541 A; 624 C; 554 G; 408 T; 0 other;

Alignment Scores:

Pred. No.: 0.000324 Length: 2127  
 Score: 146.00 Matches: 96  
 Percent Similarity: 36.36% Conservative: 64  
 Best Local Similarity: 21.82% Mismatches: 156  
 Query Match: 7.13% Indels: 124  
 DB: 21 Gaps: 25

US-09-889-746-2 (1-400) x AAZ54328 (1-2127)

Oy	18	AlaValThrGlnGlnLeuValArgLysProAsnGluSerLeuProThrValGluGln	37
Db	58	GCACTCTGCTTGGCTTTGGCCGACCAACGAAAGTTTGGAAACGCTGCATATTAA	117
Oy	38	ProVal-----ValIleThrIleAspLysSer-----	46
Db	118	GATGAGCTTCTTACAAAGCGATTGGCAGCGAATAAAGCGGATTAAGCTCGTTTGC	177
Oy	47	-----GlyMetAlaLeuAlaAsnArgIleThrGlnMetProIleThrThryVal	63
Db	178	GCCACCGTGGGTACAAAATCCCGCTTCTTTGCGGAAATTCGCAATCCGTACGATC	237
Oy	64	IleTyrgLuglnGlnIleGlnGlnGlnAlaIleThrGlySerArgGlnLeuAlaAspAlaMet	83
Db	238	ATTACCAACACGACAGGTCTAAAGACCGGAATGTTGATACG-----TTTACCAAGTTG	288
Oy	84	AlaGlnLeuIleProSerLeuGlyValSerSerGlyThrThrSerAsnPhelGlnThr	103
Db	289	GACGACAAACGCGCGCGCTGCGGTGTAGACAAAGCAAGCAAGCGCTTCGGTTAC	348
Oy	104	MetHisGlyArgGlnVal--GlnPhelLeuAsnGlyValProIleuThryLysSerArg	122
Db	349	GCGCGCGGTTACGAATACAGCAATCAACATGACGCGCTGCCC-----	393
Oy	123	AspIleSerArgGlnLeuAsnSerIleAsn-----ProAsnGlnValAla-----	137
Db	394	-----GGCGAAGTACGACAGATATACAGGACGCTGCGCCAACTGTTCGCTTGCAC	444
Oy	138	ArgIleGluValIleSerGlyValIleThrSerIleTyrg--GlySerGlyValIleThrGlyGly	156
Db	445	CGCGTGAAGTGAATGCCGCGCGGACGAGCTGTTCGACAGCGGCGGCAAGATGGGCGCG	504
Oy	157	LeuIleAsnIleValThrySerAspLeuGln-----	167
Db	505	ATCGGATCTGAGTCCGCAACGCCGACCAAAAGCTTCCAAGTCAATGCGCGGACGAG	564
Oy	168	-----GlnGluGlnPhelGlnThrArgIleGlyValHisGlySerLysLeuSerSer	184
Db	565	TTTCGGTACGCAACAACATATTAAAGCCGAGCGGACGCTATCGGCGAGC---CTCAATTCA	621
Oy	185	GlnGly-----IleGlyTyrgLysValGlnSerValAlaGlyVal	198
Db	622	GACGCGACCGTGGCGCGCGCGGTATGGCCGACACGTCGGCGCGCTCTCGCGTCCGCC	681
Oy	199	SerGluAsnGlyAsn-----ValLeuAlaArgLeuAspValAspTyrgArg	213
Db	682	GAGAAATAACAACCGCGCGGAAACTCTTACGCGCGCGGATTTGGGACATCAACCCGAT	741
Oy	214	ThrThr--GlyGlyAlaIleAspAlaAsnGlyValArgIleAlaProGlu-----	229
Db	742	ACGGTTTGGGCGCGGCTTACTTTACAGACCAACGCGCTCTCGCGCTGCACCAACGAGCTG	801
Oy	230	ProAlaGln--ThrAspLysGlnAspSerLysSerLysSerValAsnThrAsnValAsp	248
Db	802	CTTGCCGATGCCAATTAACAATTACCGTCTCTGCGGACACGATATTGTGCGCGCGAT	861
Oy	249	TyrGln-----LeuAspAspLysGlnAsnIleAsnLeuAlaLeuThryIleTyrg-----	264
Db	862	TGGAAACAATTTAAATATSCACAGCAGCAGAGTGTCGCGCATTTGAACAATTATCTTGGC	921
Oy	265	-----AsnAspLysGlnAspThrAspTyrg	272
Db	922	AACGGCGGCTACGCGCAAAAGTCGATATCGCTATTCGATCCGAAACCGATTCGAATTAT	981
Oy	273	AlaProAspTyrgLysAsnArgLeuAlaValLeuPhelGlyLysProSerLeuAsn---	291
Db	982	-----ACGTTGCGGCGAGCAA-----CTCAACAAT	1006
Oy	292	-----AlaIleTyrgLysLeuSerLysSerGlnGlnProLysThrThryLysSer	307
Db	1009	ACCGGACAAAGCGGACGCGGTTTGGTGAACGCACTTAAACAATAAGCC-----TTT	1066
Oy	308	ThrIleAsnIleLeuTyrgHisIleAspAspLeuTyrGlyValAsnThrIleAsn-----	324

Db 1063 GCGGTTACGCAAGTTACAGCCGTCCGTTTGCTGGGCAACACCCGCAAGCAATTGTG 1122  
 Oy 325 ThrAsnAlaTyrTyrArgArg-----GluIysGlyArgPheTyrProPhe 339  
 Db 1123 ATGTGTCAGACATCAACCGCTTGCGCAGTACTAATGAACAAAGGCGGT----- 1170  
 Oy 340 ValAlaProPheSerIleAlaIylAlaLeuProIleLeuGlnSerMetCAsnLeuProSer 359  
 Db 1171 -----TTCACCTTTGTCAAAAGC 1188  
 Oy 360 AlaThrLeuAspAlaTyrThrTyrValaProGlnAlaArgAlaTyrGlyValLeuGlnSer 379  
 Db 1189 GTGCGCTTAAGTATGGTTTCCGCGCTTGCTTAT-----AACGGCATACTTCAGAAC 1239  
 RESULT 27  
 ID AAA81476  
 AC AAA81476;  
 DT 04-DEC-2000 (first entry)  
 DE N. meningitidis partial DNA sequence gnm\_24 SEQ ID NO:24.  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 KW antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 KW Meningococcus B; MenB; de.  
 OS Neisseria meningitidis.  
 XX WO200022430-A2.  
 FN 20-APR-2000.  
 PD 08-OCT-1999; 99MO-US23573.  
 PF 09-OCT-1998; 98US-0103794.  
 PR 30-APR-1999; 99US-0132068.  
 XX (CHIR ) CHIRON CORP.  
 PA Frizer CM, Hickey E, Peterson J, Tettelin H, Venter JC,  
 PI Masingant V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V,  
 PI Rappunli R, Piza M;  
 DR WPI; 2000-318079/27.  
 PS Claim 7; Page 507-524; 1760pp; English.  
 CC The present invention describes methods of obtaining immunogenic  
 CC proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
 CC represent specifically claimed Neisseria meningitidis genomic DNA  
 CC sequences. AAA81260 to AAA81303 and AAB25620 to AAB25663 represent  
 CC Neisseria DNA sequences and their corresponding proteins. AAA81254 to  
 CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 CC isolation of Neisseria meningitidis DNA sequences; and AAA81322 to  
 CC AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF  
 CC sequences, which are all used in the exemplification of the present  
 CC invention. The nucleic acid sequences, protein sequences, and antibodies  
 CC against them, can be used in the manufacture of a composition. The  
 CC composition can be used as a medicament (or in the manufacture of a  
 CC medicament) for treating, preventing or diagnosing infection due to  
 CC Neisserial bacteria. For example, some of the identified proteins could  
 CC be components of vaccines against Meningococcus B; against all serotypes;  
 CC and/or against all pathogenic Neisseriae. Identification of sequences  
 CC from the bacterium will also facilitate production of biological probes,  
 CC particularly organism-specific probes. Attempts to make efficacious  
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.

CC Multivalent vaccines have also been tried but none have successfully overcome antigenic variability. The provision of further, complete sequences may provide an opportunity to identify secreted or surface exposed proteins that may be presumed targets for the immune system and CC which are not antigenically variable or at least more conserved than CC other more variable regions.

XX SQ Sequence 56485 BP; 12504 A; 14247 C; 16158 G; 13573 T; 3 other;

#### Alignment Scores:

Pred. No.: 0, 028 Length: 56485  
Score: 146.00 Matches: 102  
Percent Similarity: 36.13% Conservative: 66  
Best Local Similarity: 21.94% Mismatches: 163  
Query Match: 7.13% Indels: 135  
DB: 21 Gaps: 26

US-09-889-746-2 (1-400) x AA81476 (1-56485)

Qy 3 HisSerHisTyrPheGlnTrpLeuSerLeu----- 12  
Db 14789 CATTATCATTTATGTAATGGACAGATTATATGTCAGTTTTCGGCATCAATATGACCGCG 14848  
Qy 13 ProLeuLeuSerValAlaValThrGlnGlnLeuTyrAlaGlnProAsnGluSerLeuPro 32  
Db 14849 CCACG-GTTTTCGACACTCTCGTCTTCGGTTTTCGGCCCAACACGGAAGGTTTGGAA 14907  
Qy 33 ThrValGluLeuGluProVal-----ValIleThrIleAspLysSer----- 46  
Db 14908 ACCGTCATATTAAAGGTGACGGTCTTTACACCGCATTTGCCACCGAGAAACACGGCGAT 14967  
Qy 47 -----GlyMetAlaLeuAlaAsnArgIleThrGlnMetPro 58  
Db 14968 TACAGCTGTTTGGCGCCACCGTCGGTACAAAAATCCCGCTTCTTTGGCGAAATTCG 15027  
Qy 59 HisThrThrLysValIleTyrGluGluGlnIleGlnGluAlaThrGlySerArgGln 78  
Db 15028 CAATCCGTGAGCATATACCAACACGAGGTCAAGACCGCAATGTTGATACG----- 15081  
Qy 79 LeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSerSerGlyThrThrSer 98  
Db 15082 ---TTTGACCAAGTTGGCAGCAAAACCGCCGCTGCGGTGTGAGCAACGACGACGGA 15138  
Qy 99 AsnPheGlyGlnThrMetHisGlyArgGlnVal---GlnPheLeuLeuAsnGlyValPro 117  
Db 15139 CGCTCTTCGGTTTACGCGCGCGGTTACGAATACAGCGAATACAAATCAGCGCGCTGCC 15198  
Qy 118 LeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAsn-----ProAsn 134  
Db 15199 -----GGCCAGATGCGAGAGTATCAACGGCAGCGTGCCTCAAC 15234  
Qy 135 GlnValAla-----ArgIleGluValLeuSerGlyAlaThrSerIleTyr---GlySer 151  
Db 15235 CTGTTCCGCTTCGACCGCGTGGAGTGTGCGCGGCGGAGCGGAGTGTTCGACGACG 15294  
Qy 152 GlyAlaThrGlyGlyLeuIleAsnIleValThrLysSerAspLeuGlu----- 167  
Db 15295 GCGGAGATGGCGGCGATCTGTAATCTGTGCGCAAAACCGCCGCAACGAGGTTCCAAAGT 15354  
Qy 168 -----GluGluGlnPheGluThrArgIleGlyValHisGly 179  
Db 15355 CATGGCGCGCAGGTTTCGGTACGACAAACAATATAAAGCCGACGCGGACATCGCGGC 15414  
Qy 180 SerLysLeuSerSerGluGly-----IleGlyTyrGlnValGlyGln 193  
Db 15415 AGC---CTCAATTCAGACCGCGGTGCGCGCGGTGATGGCGCAGACCGTTCGCGCG 15471  
Qy 194 SerValAlaGlyValSerGluAsnGlyAsn-----ValLeuAlaArgLeu 208  
Db 15472 TCTCCGCGTCCCGCGAGAAACAAACCGCGCGCAACCTTCTACGCGCGCGGATGCG 15531  
Qy 209 AspValAspTyrArgThrThr---GlyGlyAlaPheAspAlaAsnGlyLysArgIleAla 227

Db 15532 GACATCAACCCCGATACGGTTTTGGCGCGCGGTATCTTTTACCAGCAACGCGCGCTCCG 15591  
Qy 228 ProGlu-----ProAlaGln---ThrAspLysGlnAspSerLysSerLeuSerVal 243  
Db 15592 CGGTACAACGCGCTGCTCGCGATGCCAATAACAAATACCGTCCCTCGCGCAACACGTA 15651  
Qy 244 AsnThrAsnValAspTrpGln-----LeuAspAspLysGlnAsnIleAsnLeuAlaLeu 261  
Db 15652 TTTGTCGCGCGGATTGGAAACAATTTAAATGCACGCCGACGTTGTTCCGCGATTG 15711  
Qy 262 ThrHisTyr-----AsnAspLys 267  
Db 15712 AAACATTTACTTCGCAACGCGGCTACGGCAAGTCGGTATGCGCTATTCGATCGGAAA 15771  
Qy 268 GlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeuPheGlyGluLys 287  
Db 15772 GCCGATTCCAATTAT-----ACGTTTGGCGGCGACAAA 15804  
Qy 288 ProSerLeuAsn-----AlaIleLysGlyLeuSerLeuSerGluGlnPro 302  
Db 15805 -----CTCAACAATACCGGACCAAGCCGACGTCAGCGGTTTGGGTACGACATTAACAA 15858  
Qy 303 LysThrThrLysSerThrPheAsnIleAsnTyrHisHisAspAspLeuTyrGlyAsnThr 322  
Db 15859 AAGGCC-----TTTGGCGTTGACGCAAGTTACAGCCGTCGCTTGGCTTGGCNAACCC 15912  
Qy 323 IleAsn-----ThrAsnAlaTyrTyrArgArg-----GluLysGly 334  
Db 15913 GCCAAGCAATTTGTGTTGTTGTCGAGACTACAAACCGCTTGGCGAGTACTAATGAACAAGG 15972  
Qy 335 ArgPheTyrProPheValAlaProPheSerIleAlaLysAlaLeuProIleLeuGlnSer 354  
Db 15973 CGT-----TCG 15978  
Qy 355 MetAsnLeuProSerAlaThrLeuAspAlaTyrThrLysAlaProGlnAlaArgAlaTyr 374  
Db 15979 ACTTGTCAAAGCGTCGCTTTAGATGTTTCCGCGCTTTCCTTAT-----AAC 16029  
Qy 375 GlyValLeuGlnSer 379  
Db 16030 GGCATACCTCAGAAC 16044  
RESULT 28  
AAF21612/C  
ID AAF21612 standard; DNA; 349980 BP.  
XX  
AC AAF21612;  
XX  
DT 13-MAR-2001 (first entry)  
XX  
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:113.  
XX  
KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;  
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
XX  
OS Neisseria meningitidis.  
XX  
PN WO200066791-A1.  
XX  
PD 09-NOV-2000.  
XX  
PF 08-MAR-2000; 2000WO-US05928.  
XX  
PR 30-APR-1999; 99US-0132068.  
PR 08-OCT-1999; 99WO-US23573.  
PR 28-FEB-2000; 2000GB-0004695.  
XX  
PA (CHIR) CHIRON CORP.  
XX  
PI (GENO-) INST GENOMIC RES.  
PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Maignani V,  
PI Galeotti C, Mora M, Ratti G, Scarfelli M, Scarlato V, Rappuoli R;

PI Frazer CM, Grandi G;  
 XX WPI; 2000-647603/62.  
 XX  
 XX Neisseria meningitidis B full length genome sequence and open reading  
 PT frames are used to detect, treat and prevent Neisserial infections -  
 XX  
 XX Claim 7; Appendix A; 692pp; English.  
 XX  
 CC The present invention describes the full length genome of  
 CC Neisseria meningitidis B (NMB). The sequences in AAP21544 and AAP21607  
 CC to AAP21613 represent fragments of the NMB genomic sequence, as the  
 CC sequence was too long to go in a record on its own it was split into 8  
 CC sequences which overlap each other at the beginning and end of each  
 CC sequence by 49980 bp (i.e. the last 49980 bp of AAP21544 is repeated at  
 CC the beginning of AAP21607, the last 49980 bp of AAP21607 are repeated at  
 CC the beginning of AAP21608, and so on). AAP21545 to AAP21588 encode the  
 CC Neisseria proteins given in AAB58550 to AAB58593, and AAP21589 to  
 CC AAP21606 represent PCR primers which are used in the exemplification of  
 CC the present invention. The NMB genome and fragments from it have  
 CC antibacterial activity, and can be used in vaccines and gene therapy.  
 CC Neisseria nucleic acids, proteins and/or antibodies which binds to the  
 CC proteins can be used in compositions for treating or preventing infection  
 CC due to Neisserial bacteria or as a diagnostic reagent for detecting the  
 CC presence of Neisserial bacteria or of antibodies raised to Neisserial  
 CC bacteria. Computers, computer memory, computer storage medium or computer  
 CC databases can be used in a search to identify open reading frames (ORFs)  
 CC or coding sequences within the NMB genome. The DNA sequences provide  
 CC further opportunities to find antigenic or immunogenic proteins which are  
 CC more effective in vaccines than the outer membrane proteins currently  
 CC used.  
 XX  
 XX SQ Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 other;

Alignment Scores:  
 Pred. No.: 0.334 Length: 349980  
 Score: 146.00 Matches: 102  
 Percent Similarity: 36.13% Conservative: 66  
 Best Local Similarity: 21.94% Mismatches: 163  
 Query Match: 7.13% Indels: 135  
 DB: 21 Gaps: 26

US-09-889-746-2 (1-400) x AAP21612 (1-349980)  
 QY 3 HisSerHisrPhegIntRpleuSerLeu----- 12  
 DB 126865 CATTTATCATTTATGTGATGAGACAGTTATGTCAGTTTCCGATCATATATACCGCCG 126806  
 QY 13 ProLeuLeuSerValAlaValThrGlnGlnLeuTyzAlaGlnProAsnGlnSerLeuPro 32  
 DB 126805 CCAAG-GTTTGGCAGCAGCTCGTCTGCGTTTGGCCGACAAACGAAAGGTTTGGAA 126747  
 QY 33 ThrValGlnLeuGlnProVal-----ValIleThrIleAspIleSer----- 46  
 DB 126746 ACCGTCATATTAGAGGTCAGCGTTCTTACACGCGATTCACCGGAAACACGCGCAT 126687  
 QY 47 -----GlyMetAlaLeuAlaAsnAfgIleThrGlnMetPro 58  
 DB 126686 TACAGCTCGTTGGCCGACCGTCGCTGATACAAATAATCCCGCTTGTGGCCGAAATTCGG 126627  
 QY 59 HisThrThrIleValIleTyrgIuGlnGlnIleGlnGlnAlaThrIleSerArgGln 78  
 DB 126626 CAATCGTCAGCATCATATACCAACGACGATGCAAGACGCAATGTATAG----- 126573  
 QY 79 LeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSerSerGlyThrThrSer 98  
 DB 126572 ---TTTGACAGGTGGACCAAAACCGCCGCTGCGCTGTGACCAACGACGACGA 126516  
 QY 99 AsnPheGlyGlnThrMetHisGlyArgGlnVal---GlnPheLeuLeuAsnGlyValPro 117  
 DB 126515 CGCTCTTCGGTTTACCGCGCGGCTTACGAATACAGCAATACCAACATCCGCTGCGCC 126456  
 QY 118 LeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAsn-----ProAsn 134

DB 126455 -----CGCGAGATGCAGAGATATCAACGCGACGCTGCCAAC 126420  
 QY 135 GlnValAla-----ArgIleGluValLeuSerGlyAlaThrSerIleTyr---GlySer 151  
 DB 126419 CTGTTGCGCTTTCGACCGCGGATGATGCGCGGCGCGGACGATGTCAGACAGC 126360  
 QY 152 GlyAlaThrGlyGlyLeuIleAsnIleValThrIleSerAspLeuGlu----- 167  
 DB 126359 GCGGAGATGCGCGGATGATGATCTGGTGGCCCAACCGCCGACCAAGCGTTCCAGGT 126300  
 QY 168 -----GluGlnGlnPheGluThrArgIleGlyValHisGly 179  
 DB 126299 CATGCGCGCGCAGGTTGCGTACGACCAAAACATATTAAGCCGACGCGACGTATCGGCG 126240  
 QY 180 SerIleLeuSerSerGlnGly-----IleGlyTyrgIuValGlyGln 193  
 DB 126239 AGC---CTCAATTCAGACCGCGCGCTGCGCGCGCGCGTATGCGCAACCGTGGCGCG 126183  
 QY 194 SerValAlaGlyValSerGluAsnGlyAsn-----ValLeuAlaArgLeu 208  
 DB 126182 TCTCCGCGTCCCGCGGAGAAACACCGCGCGAAACCTTCTACGCGCGCGGATTTGG 126123  
 QY 209 AspValAspTyrgIuThrThr---GlyGlyAlaPheAspAlaAsnGlyValArgIleAla 227  
 DB 126122 GACATCAACCCGATACGATGTTTGGCGCGGCGCTATCTTACCAACGACGCGCTCGCG 126063  
 QY 228 ProGlu-----ProAlaGln---ThrAspIleGlnAspSerIleSerLeuSerVal 243  
 DB 126062 CGGTACACAGCGCTCGCTGCGCGATGCCAATTAACCTTCCTCGCGCAACAGCTA 126003  
 QY 244 AsnThrAsnValAspTyrgIuGln-----LeuAspAspIleGlnAsnIleAsnLeuAlaLeu 261  
 DB 126002 TTGTCGCGCGCGATTTGACAAACATTTAAATGACACGACGACGATGTCGCCGATTTG 125943  
 QY 262 ThrHisrTyT-----AsnAspIle 267  
 DB 125942 AAACATTAATTCGCGCAACGCGGCTACGCGCAAGTCGATGCGCTAATTCGATCGGAAA 125883  
 QY 268 GlnAspThrAspTyrgIuAlaProAspTyrgIuAsnAlaGlyLeuAlaValLeuPheGlyGlyIle 287  
 DB 125882 GCCGATTCGAATTAAT-----ACGTTTGGCGCGCGCAACAA 125850  
 QY 288 ProSerLeuAsn-----AlaIleTyrgIleuSerLeuSerGlnGlnPro 302  
 DB 125849 -----CTAACCAATACCGCAACGCGGATGCGGTTTGGTACGACATTAACAA 125796  
 QY 303 TyThrThrIleSerThrPheAsnIleAsnTyrgIuHisAspAspLeuTyrgIuAsnThr 322  
 DB 125795 AAAGCC-----TTGCGGTTGACGCAAGTTACAGCCGTCGTTGCTGGGCAACACC 125742  
 QY 323 IleAsn-----ThrAsnAlaTyrgIuArgG---GluTyrgIle 334  
 DB 125741 GCCACGAATTTGTGATGTTGTCAGCTACACCGCTTGGCAGTACTAATTAACAGAG 125682  
 QY 335 ArgPheTyrgIuPheValAlaProPheSerIleAlaTyrgIuLeuProIleLeuGlnSer 354  
 DB 125681 CGT-----TCG 125676  
 QY 355 MetAsnLeuProSerAlaThrIleAspAlaTyrgIuValProGlnAlaArgAlaTyrg 374  
 DB 125675 ACTTGTCAAAAAGCGTGCCTTGAATGTTTCCGCGCTTGCCTTAT-----AAC 125625  
 QY 375 GlyValLeuGlnSer 379  
 DB 125624 GGCATATCTTCAGAAC 125610  
 RESULT 29  
 ID AAA81489/c  
 AC AAA81489 standard; DNA; 837096 BP.  
 XX  
 XX AAA81489;



```

QY 303 IyeThrThlySerThrPheAsnIleAsnTyHHisAspAspLeuTrpGlyAsnThr 322
    |||
Db 490566 AAAGCC-----TTTCGGTTGACGCAAGTTACAGCGCTCGTTGGCTTGGCAACACC 490513
    |||
QY 323 IleAsn-----ThrAsnAlaTyTrpArgArg-----GluysGly 334
    |||
Db 490512 GCCACGAAATTGTGATGGTGCAGACTACACCGCTTGGCAGACTACTAATGAAACAGGG 490453
    |||
QY 335 ArgPheTy-PropheValAlaProPheSerIleAlaLysAlaLeuProIleLeuGlnSer 354
    |||
Db 490452 CGT-----TCG 490447
    |||
QY 355 MetAsnLeuProSerAlaThrLeuAspAlaTyTrThlyAsnAlaProGlnAlaArgAlaTy 374
    |||
Db 490446 ACTTTGTCAAAAGCGTCGCTTGAATGGTTCCGCGCTTGCCTTAT-----AAC 490396
    |||
QY 375 GlyValLeuGlnSer 379
    |||
Db 490395 GGCATACTTCAGAAC 490381
    |||

```

## RESULT 30

ID AAAS9217 standard, DNA; 2078 BP.

AC AAAS9217;

DT 07-NOV-2000 (first entry)

DE DNA encoding a *Neisseria meningitidis* BASB053 polypeptide.

KW BASB053; *Neisseria meningitidis* infection; vaccine; ss.

OS *Neisseria meningitidis*.

```

FH Key      Location/Qualifiers
FT CDS      2..2078
FT          /tag=a
FT          /product="BASB053"

```

PN MO200042193-A1.

PD 20-JUL-2000.

PF 10-JAN-2000; 2000MO-BP00137.

PR 15-JAN-1999; 99GB-0000959.

PR 28-JAN-1999; 99GB-0001903.

XX (SMRK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Ruelle J;

DR WPI; 2000-476062/41.

DR P-PSDB; AAB07698.

PT New *Neisseria meningitidis* polypeptide useful for diagnosis of  
 CC *Neisseria* infection and for development of vaccines against such  
 CC infection -

XX Claim 11; Page 56; 92BP; English.

XX The present sequence encodes a *Neisseria meningitidis* BASB053  
 CC polypeptide. The BASB053 polypeptide, or an antibody immunospecific  
 CC for BASB053, may be identified in a biological sample in order to  
 CC diagnose a *Neisseria meningitidis* infection in an animal. The BASB053  
 CC polypeptides and polynucleotides may be used as vaccines, for  
 CC generating an immune response in an animal. A composition comprising  
 CC at least one antibody immunospecific for BASB053 may be used to  
 CC treat humans infected with *Neisseria meningitidis*.

XX Sequence 2078 BP; 523 A; 639 C; 546 G; 370 T; 0 other;

```

Alignment Scores:
Pred. No.: 0.000393
Score: 145.00
Percent Similarity: 37.77%
Best Local Similarity: 22.76%
Query Match: 7,084
DB: 21
Gaps: 22

```

US-09-889-746-2 (1-400) x AAAS9217 (1-2078)

```

QY 47 GlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrThlyValIleTyGlu 66
    |||
Db 96 GGCACAAAATCCCGCTTCTTTCGCCGGAATTCGCATCGCATGATCATCAACCAAC 155
    |||
QY 67 GlnGlnIleGlnGlnGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeu 86
    |||
Db 156 CAGCAGGTCAAAAGCCGCAATGTGATACG-----TTTGACCAAGTTGGCGCCAAA 206
    |||
QY 87 IleProSerLeuGlyValSerSerGlyThrThrSerAsnPheGlyGlnThrMetHisGly 106
    |||
Db 207 ACGCCCGCGCTGCGGTGTGAGCAACATGACGACGCGCTTCGGTTACGGCGCGGT 266
    |||
QY 107 ArgGlnVal--GlnPheLeuLeuAsnGlyValProLeuThrGlySerArgAspIleSer 125
    |||
Db 267 TACGAATACAGCAATACAAATCAGCGCTCGCC----- 302
    |||
QY 126 ArgGlnLeuAsnSerIleAsn-----ProAsnGlnValAla-----ArgIleGlu 140
    |||
Db 303 GCGCAGATCAGAGATATCAACGCGACGCTGCCCAATCTGTTCGCTTCACCGCGTGA 362
    |||
QY 141 ValLeuSerGlyAlaThrSerIleTy--GlySerGlyAlaThrGlyValIleuIleAsn 159
    |||
Db 363 GTGATGCGCGGCGCGGCGGAGCTGTTCAGACAGCGGGAATGGCGGTATCGTAAAT 422
    |||
QY 160 IleValThrThlySerAspLeuGlu----- 167
    |||
Db 423 CTGGTGGCGCAACCGCCGCAAGAGCTTCAAGATCATGTCGCGCAGGCTTCGATACG 482
    |||
QY 168 GlnGlnGlnPheGlnThrArgIleGlyValHisGlySerLysLeuSerSerGlyGlu 186
    |||
Db 483 CACAAACAAATATTAACCGCAGGCGGACGATTCGCGAGC--CTCAATTCAGACGGCAGC 539
    |||
QY 187 -----IleGlyTyrglnValGlyGlnSerValAlaGlyValSerGlnAsn 201
    |||
Db 540 GTGCGCGCGCGGTATGCGCAGACCGTCCGCGGTCTCCGCGTCCCGCGCAAAAAC 599
    |||
QY 202 GlyAsn-----ValLeuAlaArgLeuAspValAspTyArgThr 215
    |||
Db 600 AACCGCAGCAAAACCTTCTACCGCGCGGAGTTGGACATCAACCCGATCGATTGG 659
    |||
QY 216 GlyGlyAlaPheAspAlaAsnGlyLysArgIleAlaProGlu-----ProAlaGln 232
    |||
Db 660 GCGCGCGGCTATCTTACCGACGACGCACTCGCGCGGTACACGCGCTTGCACGCGAT 719
    |||
QY 233 --ThrAspLysGlnAspSerLysSerLeuSerValAsnThrAsnValAspTyrgln 250
    |||
Db 720 GCCAATTAACAATTCACGCTCCCGCGCAACAGTATTGTGCGCGCGGATTGGAACAA 779
    |||
QY 251 --LeuAspAspLysGlnAsnIleAsnLeuAlaLeuThrHis----- 263
    |||
Db 780 TTAAATATGAAACAGCACACGCTGTTGGCGATTGAAACATTACTTCGCAACGCGCG 839
    |||
QY 264 -----TyrAsnAspLysGln--AspThrAspTyrgln----- 273
    |||
Db 840 TACGGCAAAAGTCGTATGCGCTATTTCACCGCGATGCGCATCAACATATCCCTTGGCC 899
    |||
QY 274 -----ProAspTyrglyAsnArgLeuAlaValIleLeuPheGlyG1 286
    |||
Db 900 GCGAGCAAGCTGGCATGAACACCCCGGAGCGCCCGCGGTGCAATACGCTACAGAC 959
    |||
QY 286 uLySPro----- 288
    |||
Db 960 AAAGCTGCGCGGTGGTTGGGTACAGAAATCAAAACAAAGACCTCGCTTTCAGCGCC 1019
    |||

```

QY 289 -----SerLeuAsnAlaIlelysglyLeuSerLeuSerGluGlnProly 303  
 Db 1020 AGCTACAGAGCGCTTTCGCTGGGCAATACGGCAACAAATTTGTCTATCGCGCCGAT 1079  
 QY 303 sThrThrlySerThrPheAsnIleAsn-----TyrHisHisAspAsePle 318  
 Db 1080 TACACCGCTTCGGAGCACCACGACGAGCGCTACTTATATATACGCGCGCGC 1139  
 QY 318 uTrpGlyAsnThrIleAsnThrAsnAlaTyrTyrArgArgGluLysGlyArgPheTyrPr 338  
 Db 1140 CTGGCTTTAAACG-----AGTTCGCAGCATACGCGAGG----- 1173  
 QY 338 oPheValAlaProPheSerIleAlaLysAlaLeuProIleLeuGlnSerMetAsnLeuPr 358  
 Db 1174 -TTGATTTGATTCACCAACGCGCAAGCGTGCAGCGGTACAGCCATACCGTCTAC 1232  
 QY 358 o--SerAlaThrLeuAspAlaTyrThrLysAlaPro 369  
 Db 1233 GAAACCTCGACGAATTCGCGCATTTACGGCAATCCA 1269

RESULT 31  
 ID AAS82072  
 XX AAS82072 standard; cDNA; 679 BP.  
 AC AAS82072;  
 XX 13-FEB-2002 (first entry)  
 XX DNA encoding novel human diagnostic protein #17876.  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 XX Homo sapiens.  
 XX WO200175067-A2.  
 XX 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US08631.  
 XX 31-MAR-2000; 2000US-0540217.  
 XX 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 XX WPI; 2001-639362/73.  
 XX P-P8DB; ABG17885.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX Claim 1; SEQ ID No 17876; 103pp; English.  
 XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS84197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 679 BP; 168 A; 172 C; 181 G; 158 T; 0 other;

Alignment Scores:  
 Pred. No.: 9.59e-05 Length: 679  
 Score: 144.50 Matches: 68  
 Percent Similarity: 36.36% Conservative: 36  
 Best Local Similarity: 23.78% Mismatches: 87  
 Query Match: 7.06% Indels: 95  
 DB: 23 Gaps: 15

US-09-889-746-2 (1-400) x AAS82072 (1-679)

QY 113 LeuAsnGlyValProLeuThrGlySer-----ArgAspIleSerArgGlnLeuAsnSer 130  
 Db 85 CTGAATGGCTGAAGTTGCAGGCAACTTCTATAACGATCG-----GTC 129  
 QY 131 IleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyAlaThrSerIle----Tyr 149  
 Db 130 ATTGACCCGTATATGCTGGAACGCGCTGAATATATGCTGCCCGGTTTCGTGCTTAC 189  
 QY 150 GlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLys-----SerAsp 165  
 Db 150 GGTAAACCAATCTCGCGCGCTGTGTGANTATGTTCAGCAAGCTCCGACCAACCGG 249  
 QY 166 LeuGluGluGlnPheGluThrArgIleGlyValHisGlySerLysLeuSerGlu 185  
 Db 250 CTGAAAGAAAGTTCAGTTT-----AAAGCCGGTACTGCAC 282  
 QY 186 GlyIleGlyTyrGlnValGlyGlnSerValAlaGlyValSerGluAsnGlyAsnValLeu 205  
 Db 283 AGCCTG---TTCCAGACTGGTGTGACTTTAGCGATTCGTTGGATGATGACGGTGTTTAC 339  
 QY 206 AlaArgLeuAspValAspTyrArgThrThrGlyGlyAlaPheAspAlaAsnGlyLysArg 225  
 Db 340 TCT-----TATCGCCTGACCGCTCTGGCGCTCTGCGCAAT----- 375  
 QY 226 IleAlaProGluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsnThr 245  
 Db 376 -----GCCAGCAGAAAGGTCAGAAAGCAGCAGCGTTATGCTATTGCACCG 420  
 QY 246 AsnValAspTyrGlnLeuAspAspLysGlnAsnIleAsnLeuAlaLeuThrHisTyrAsn 265  
 Db 421 CGCTTCACCTGGCGTCCGATGATAAAACCAATTTTACCTTC---CTTTCTTACTCCAG 477  
 QY 266 AspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeuPheGly 285  
 Db 478 AACGAGCCGGATACCGGTTAT-----TACGCG 504  
 QY 286 GluLysProSerLeuAsnAlaIleLysGlyLeuSerLeuSerGluGlnProLysThrThr 305  
 Db 505 TGGTTGCCGAAAGAGGAAACCGTTGAGCGCGCTG-----CCGAAACGGTAAGCGTTTG 555  
 QY 306 LysSerThrPheAsnIleAsnTyrHisHisAspAspLeuTrpGlyAsnThrIleAsnThr 325  
 Db 556 CCGACAGACTTTAAAT-----GATAGCGCGAAGAAC 585  
 QY 326 AsnAlaTyrTyrArgArgGluLysGlyArgPheTyrProPheValAlaProPheSerIle 345  
 Db 586 AACGCTTATTCGTAAATGAGAAG----- 609  
 QY 346 AlaLysAlaLeuProIleLeuGlnSerMetAsnLeuProSerAlaThrLeuAspAlaTyr 365  
 Db 610 -----ATGAAC----- 615  
 QY 366 ThrLysAlaProGlnAlaArgAlaTyrGlyValLeuGlnSerGluSerLysAlaGluVal 385



Dd		616 -----GAAACCGAAGCAGGTATTATTCGGCTTGCAGGAAGCAGAACCGTTGAACCG 669          :::
Oy	386 LeuGlYarGVaIProAn 391     	
Dd	670 CTG-----CCGAAC 678 	
RESULT 32		
ID	AZ36926/c	
XX	AZ36926 standard; DNA; 8907 BP.	
XX	AZ36926;	
XX	13-MAR-2000 (first entry)	
DE	Nucleotide sequence of the genome of Stealth virus clone 313.	
XX	Prototype Stealth virus clone; atypically structured virus; KM vacuolating cytopathic effect; cytopathic virus; virus detection; KM malignancy; multiple myeloma; lymphoma; brain tumour; breast cancer; KM salivary gland tumour; Alzheimer's disease; Parkinson's disease; KM spongiform encephalopathy; multiple sclerosis; schizophrenia; KM Rett's depression; major depression; personality disorder; autism; KM Rett's syndrome; attention deficit; oppositional defiance; aggression; anorexia nervosa; bulimia; multi-system illness; virus infection; ss. XX	
OS	Stealth virus.	
XX		
PN	MO9960101-A1.	
PD	25-NOV-1999.	
XX		
Ff	19-MAY-1999; 99WO-US11185.	
XX		
PR	19-MAY-1998; 98US-0081708.	
XX	(MART/) MARTIN W J.	
PA	Martin WJ;	
XI		
DR	WPI: 2000-072436/06.	
XX		
PT	Isolated viruses, used to develop products for detection, prevention and treatment of stealth virus infections -	
PS	Disclosure; Page 34-37; 50pp; English.	
XX		
CC	AZ36925-30 represent the nucleotide sequences of prototype Stealth virus clones. The sequences illustrate the type of sequences that can be used to generate probes and to predict antigenic and biologically active products obtainable from a viral isolate, in the methods of the invention. The specification describes tissue culture, serological and molecular based methods to detect atypically structured viruses, such as Stealth viruses, which are able to induce a vacuolating cytopathic effect (CPE) in tissue culture, and are distinguishable from known cytopathic viruses. By appearance progression and/or host range, or by serological, electron microscopic and/or molecular markers. The products can be used for detecting viruses in patients with diseases such as a malignancy, e.g. multiple myeloma, lymphoma, brain tumours, breast cancer, salivary gland tumours, Alzheimer's disease, Parkinson's disease, spongiform encephalopathy, multiple sclerosis, schizophrenia, manic depression, major depression, personality disorders, autism, Rett's syndrome, attention deficit, oppositional defiance, aggression, anorexia nervosa, bulimia, a multi-system illness, an animal illness or an illness in a domestic dog or cat. The products can be used for detecting Stealth viruses in biological products such as blood products and foods. The products can also be used for developing agents for treating or preventing virus infections. XX	
SQ	Sequence 8907 BP; 2257 A; 2447 C; 2165 G; 2023 T; 15 other;	
Alignment Scores:		
Pred. No.:	0.00396	Length: 8907

Score:	143.50	Matches:	97
Percent Similarity:	40.34%	Conservative:	68
Best Local Similarity:	23.72%	Mismatch:	126
Query Match:	7.01%	Indels:	119
DB:	21	Gaps:	23

  

US-09-889-746-2 (1-400) x AAZ63926 (1-8907)	
OY 10 LeuserieProleuleSerValAlaValThrGlnLeuTyrAlaGlnProAsnGlu 29	
Db 1758 GTGGCGGCGCGTTGGATGTCTATCCGACAGACAC---AACTTAAGACAGAGATTGC-- 17050	
OY 30 SerieuprothrValGluLeugInuproValIleThrIleAspIysSerGlyMetAla 49	
Db 1704 -----GTCGTGCTTGACAGATGTTGTAGC-----CCGTGCC 16699	
OY 50 LeuAlaAsnArgIleThrgInMePProIsthTrHlysValIleTyrGluGluGlnIle 69	
Db 1668 CGCGGCAACATCGTTGACGCGCTTCAAGCTCTCCGTCACGCTGATTAAGAAGCCGAGATT 16099	
OY 70 GlnGluGlnAlaThrGlySerArGlnLeuAlaAspValMetAlaGlnLeuIleProser 89	
Db 1608 GAGCGCTCTGCTTCACCGCATCTCAA-----TCGCTCTCCAATAATTATCCGGC 15588	
OY 90 LeuGlyValSerSer-----GlyThrThrSerAsn---PheGlyGlnThrMet 104	
Db 1557 ATTTTCGATCACCGGAATGCGCGACAGAGATGCTCTTCAACTCTAATTATTCGGGGCATG 14989	
OY 105 HisGlyArgGlnValGlnPheLeuLeuAsnGlyValProLeuThryGlySerArgApile 124	
Db 1497 TCATCGAACAAGACAGATCGTTCTGTGGAAATGGCGTCCGACAGCTTCGGCAGCACCGGG 14388	
OY 125 SerArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGly 144	
Db 1437 GCAGCTGCGCTTCCAAATATTCGTTGAGCTCAATTGATGATGATGAGATCGCCAAAGC 13787	
OY 145 Ala---ThrSerIleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrIys 163	
Db 1377 GCGCATTCCTCGCAATACGGTGGCGATGGCGGAGTGGCGGTATCAACATTAACCAAG 13118	
OY 164 SerAsp-----LeuGluGluGluGlnPhe---GluThrArgIleGly----- 176	
Db 1317 CAGGCGGAGCGTGTGGCCAGCGCTCTTCTCGGAGCGTACACAGGTGTAAAGCAC 12588	
OY 177 -----ValHisGlySer-----LysLeuSerSerGluGlyIleGlyTyr 189	
Db 1257 CCATGGGGCGGTATGTATATCGGCTCTCTGCAAGGCGGACGAAGAAGATGATCATATT 11989	
OY 190 GlnValGlyInSerValAlaGlyValSerGluAsnGlyAsnValLeuAlaArgLeuSp 209	
Db 1197 GCCTTTGTGGCGACGACCGCG---ACGCAAGGCTTAATAATTC----- 11566	
OY 210 ValAspTyrArgThrThrGlyGlyAlaPhe-----AspAlaAsnGlyIyAsArg 225	
Db 1155 -----KCGACTCCCGAAGCTTTTGTTATGACCTGTATCGGATGCT-- 11144	
OY 226 IleAlaProGluProAlaGlnThrAspIysGlnAspSerIysSerLeuSerValAsnThr 245	
Db 1113 -----TTCCGSCAGGATCTCTTAACACTTGGCGCTGTCAAG 1078	
OY 246 AsnValAspTyr-----GlnLeu 251	
Db 1077 GATTTTGTCTGGGCGACAGATTATGCGCATGTCTTCTAGCCGACGACATCAATAT 1018	
OY 252 AspAspIys---GlnAsnIleAsnLeuAla---LeuThrHisTyrAsnAspIys----- 267	
Db 1017 GATGCAAAAGCGCGCTGCGCAAGAGCTTACACAAACCGCTTAGCGGCAAGCTTGGC 958	
OY 268 GlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeuPheGly----- 285	
Db 957 GCGCGGGTCAATCACTTGAAGCTGTCTGTGACATTTGA-ATTCTGTACGGCTACTTT 899	
OY 286 -----Glu 286	



```

Db 1383 GCGCATTCCTCCGCAATACGGTGCATGCGCGGCGGCGCTCATCAACATATACCAAG 1324
Qy 164 SerAsp-----LeuGluGluGluGlnPhe-----GluThrArgIleGly----- 176
Db 1323 CAGGCGGAGCGGTGTGGCGGCGCTCTTCTGCGGAAGCGCTGACCAACAGCTGAACGAC 1264
Qy 177 -----ValHisGlySer-----LysLeuSerSerGluGlyIleGly 189
Db 1263 CCATGGGGCGGTATGATATCCGCTCTCTGCAAGGGCGGCAAGATGATTCATAT 1204
Qy 190 GlnValGlyGlnSerValAlaGlyValSerGluAsnGlyAsnValLeuAlaArgLeuAsp 209
Db 1203 GCGTTTGTCGCGCAGTGAACCGGC---ACCGAGGCGCTATATTTTC----- 1162
Qy 210 ValAspTyrArgThrGlyGlyAlaPhe-----AspAlaAsnGlyLysArg 225
Db 1161 -----ACGACTCCCGAAGCTTTGGTTATGAGCGCTATGCGCGT----- 1120
Qy 226 IleAlaProGluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsnThr 245
Db 1119 -----TTCCTGCAGGATCGTTCACTTCGCGCTGTGGAAG 1084
Qy 246 AsnValAspTyr-----GlnLeu 251
Db 1083 GATTTTGACTGGGGCAAGATTATGCGGATGCTTCTTGAGCCGAGAGCGCAATCAATAT 1024
Qy 252 AspAspLys---GlnAsnIleAsnLeuAla---LeuThrHisTyrAsnAspLys----- 267
Db 1023 GATGCAAAAGCCGCGTGGCGCAAGAGCTTACCAACAGCCCTTACGCGCAAGCTTGGC 964
Qy 268 GlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeuPheGly----- 285
Db 963 GCGCGGTGATACAGTTTCAAGCTGTCTCTGACGATTGA-ATTCTGTACGCGTACTTT 905
Qy 286 -----Glu 286
Db 904 TCTTTAATGCGTGCATCATCATATGAGATTGAGTATCGTTCCATGCGCATGAGATCAT 845
Qy 287 LysProSerLeuAsnAlaIleLysGlyLeuSerLeuSerGluGlnProLysThrLys 306
Db 844 GAGCCAGCGCTTGAACGAAACAGGGGGGTGA-----GAGCTGATGAGAACTGAT 794
Qy 307 SerThrPheAsnIle-AsnTyr----- 313
Db 793 ACATGCTTCCGATTAATGCTGCTAGTTGAGCTCATATATCATCATATCGGGTAGCG 734
Qy 314 -----HisHisAspAspLeuTyrGlyAsnThrIleAsnThrAsnAlaTyrTyrArgAr 331
Db 733 TGGCAGTAGTAACGATGTCGCTATGCTAGTATTCAGAAAGTAGACGTC-----GC 683
Qy 331 GdlLysGlyArgPheTyrProPhe 339
Db 682 TGGCAAAAGTAAAGTTTTCGCTTTT 658

RESULT 34
AAS46270/c
ID AAS46270 standard; DNA; 10353 BP.
XX
XX AAS46270;
XX
XX 18-DEC-2001 (first entry)
XX
XX DNA encoding novel mar regulated protein (NIMR) #39.
XX
XX mar regulated polypeptide; NIMR; microbial infection; antibacterial; ds.
XX
XX Escherichia coli.
XX
XX MO20017076-A2.
XX
XX 27-SEP-2001.
XX

```

```

PF 08-MAR-2001; 2001MO-US07478.
XX
XX PR 10-MAR-2000; 2000US-188362P.
XX
XX PA (TUFT ) TUFTS COLLEGE.
XX
XX PI Levy SB, Barbosa TM, Alekshun NN;
XX
XX DR WPI: 2001-602769/68.
XX
XX DR P-PSDB; MAU29371.
XX
XX
XX PS Disclosure; Page 451-457; 526pp; English.
XX
XX CC The invention relates to a method of identifying compounds that modulate
XX CC a newly identified mar regulated (NIMR) polypeptide activity. The method
XX CC comprises contacting a NIMR polypeptide with a test compound under
XX CC interaction conditions, determining the ability of the compound to
XX CC modulate the activity or expression of the polypeptide, and selecting the
XX CC modulators. NIMR nucleic acids and polypeptides are used in the treatment
XX CC of microbial infections, and in screening for modulators of NIMR
XX CC expression and activity. These modulators can be used to reduce the
XX CC infectivity of a microbe on a surface, and the virulence of a microbe in
XX CC a subject suffering from an infection. AAS46232-AAS46278 represent
XX CC Escherichia coli NIMR coding sequences of the invention.
XX
XX SQ Sequence 10353 BP; 2559 A; 2763 C; 2485 G; 2546 T; 0 other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 0.00608 Length: 10353
XX Score: 142.50 Matches: 79
XX Percent Similarity: 39.58% Conservative: 52
XX Best Local Similarity: 23.87% Mismatches: 109
XX Query Match: 6.96% Indels: 92
XX DB: Gaps: 14
XX
XX US-09-889-746-2 (1-400) x AAS46270 (1-10353)
XX
XX Qy 40 ValIleThrIleAspLysSerGly-----MetAlaLeuAlaAsn 52
XX Db 10112 GGTTAAGCGGTGATGATGATGCGAAGCATGTTGTCATGCATCTTCCTGGAGACA 10053
XX Qy 53 ArgIleThrGlnMetProHisThrLysValIleTyrGluGluGlnIleGluGln 72
XX Db 10052 AATCTTAAAGATGACCTGCCAGTATCAGCTCATTAACCCAGGAAGACCTGACGCAAA 9993
XX Qy 73 AlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyVal 92
XX Db 9992 CCGGTA-----CAGAAATTAAGGATGCTCAAGAA---GTCCCTGGCTACACAG 9942
XX Qy 93 SerSerGlyThrThrSerAspPheGlyGlnThrMetHisGlyArgGlnValGlnPheLeu 112
XX Db 9941 ACGAAGCAAGGGGATTAACCGTAAGGCGCTATATTCGGTCTGAGACAGCATATACC 9882
XX Qy 113 LeuAsnGlyValProLeuThrGlySerArgAspIleSerArg----- 126
XX Db 9881 CTG-----ATTCTGCTCAGCGTAACCGCTGAACCTCCGCAATGCGCTTCCGCCAC 9828
XX Qy 127 -----GlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSer 143
XX Db 9827 AATGATTCGATCGTAACCGATCCCGGTGCAATTCATGCAAGATTAAGTGCCT 9768
XX Qy 144 Gly---AlaThrSerIleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThr 162
XX Db 9767 GGGCCGATGCTCGCTGTACGCTTCCAGTGGCGCTGCTGATATCATCAC 9708
XX Qy 163 LysSerAspLeuGluGluGlnPheGluThrArgIleGlyValHisGlySerLysLeu 182
XX Db 9707 AAA-----AAATTC 9699
XX

```

Qy 183 SerSerGluGlyIleGlyTyrGlnVal-----GlyGlnSerValAla 196  
 Db 9698 GG-TGAGAAATGGTCGGGTACCGTTACCTCGATACCACCATTCAGGAACATCGCGATCG 9640  
 Qy 197 -GlyValSerGluAenGlyAsnValLeuAlaArgLeuAepValAepTyrArgThrGlu 216  
 Db 9639 CGGTGACACCTATACGGT-----CAGTTCTTTACCACTGG 9604  
 Qy 216 Y-----GlyAlaPheAepAlaAenGlyLysArgIleAl 227  
 Db 9603 ACCATTAAATTGATGGTGTCTGGGAATGAAGCTTACGGCAGCTGGCAAAACGTGAAAA 9544  
 Qy 227 aProGluProAlaGlnThrAspLysGlnAspSerLysSerLeu----- 241  
 Db 9543 GGATGACCGCGAAAACCTCAACGACCACTACCGGAGAAACCGCGPATTGAAGGATT 9484  
 Qy 242 ---SerValAenThrAsnValAepTyrGlnLeuAepAspLysGlnAsnIleAenLeuAl 260  
 Db 9483 CTCACCGCGGACGGCAATGTGAATTCCTCTGGACACCGCAATCAAAATCACGATTTTAC 9424  
 Qy 260 aLeuThrHisTyrAsnAspLysGlnAepThrAep----- 271  
 Db 9423 TCGCGGATACGGTTTACCGGTTCAGGATCGTATTCGGACTCGTGACAAAAACCGCCT 9364  
 Qy 272 -----TyrAlaPro-----AspTyrGlyAsnArgLeuAl 281  
 Db 9363 GGAACCGCCAGAACTACTCGCTCAGCCATATATGGCGTTACGGCACCAGCGAACT 9304  
 Qy 281 aValLeuPheGlyGluLysProSerLeuAenAlaIleLysGlyLeuSerLeuSerGluGl 301  
 Db 9303 GAAATACTACGTGAGAAAGTCAG---AACAAAAACCTTGGACACGAGCCCGATAC 9247  
 Qy 301 n-ProLysThrThrLysSerThrPheAsn 310  
 Db 9246 TTCCGAAGCAATACGTCGCGGCAAT 9218  
 RESULT 35  
 ABQ90127  
 ID ABQ90127 standard; DNA; 1914 BP.  
 XX AC ABQ90127;  
 XX DT 01-OCT-2002 (first entry)  
 XX DE M. capsulatus gene #112 for DNA array.  
 XX KW Micro array; gene; db; differential expression; gene expression.  
 XX OS Methylococcus capsulatus.  
 XX PN WO200255655-A2.  
 XX PD 18-JUL-2002.  
 XX PF 14-JAN-2002; 2002WO-NO00019.  
 XX PR 12-JAN-2001; 2001NO-0000235.  
 XX PR 12-JAN-2001; 2001NO-0000239.  
 XX XX (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.  
 XX PA (TIGR-) TIGR.  
 XX PI Birkeland NK, Eidhammer I, Jonassen I, Jensen HB, Lien T;  
 XX PI Lillehaug JR, Lossius I, Eisen JA, Fraser CM, Durkin AS;  
 XX PI Salzberg SL;  
 XX DR WPI; 2002-557818/59.  
 XX XX Novel DNA array useful for determining differential expression of  
 PT Methylococcus capsulatus genes, comprises polynucleotides or  
 PT oligonucleotides representative for a selective number of Methylococcus  
 PT capsulatus genes -

PS Claim 19; Page 104-105; 678pp; English.  
 XX The invention relates to a novel DNA array giving a representation of a  
 CC number of Methylococcus capsulatus genes. The method of the invention is  
 CC useful for determination of the differential expression of the genes of  
 CC M. capsulatus, and for studying gene expression on a genomic scale and in  
 CC gene expression assays of M. capsulatus genes. The sequences shown in  
 CC ABQ9016-ABQ91855 represent M. capsulatus genes for use in arrays of the  
 CC invention.  
 XX SQ Sequence 1914 BP; 414 A; 663 C; 530 G; 307 T; 0 other;  
 Alignment Scores:  
 Pred. NO.: 0.00107 Length: 1914  
 Score: 140.00 Matches: 93  
 Percent Similarity: 39.90% Conservat: 59  
 Best Local Similarity: 24.41% Mismatches: 140  
 Query Match: 6.84% Indels: 90  
 DB: Gaps: 22  
 US-09-889-746-2 (1-400) x ABQ90127 (1-1914)  
 Qy 3 HisSerHisTyrPheGln-----TrpLeuSerLeuProLeuLeu 15  
 Db 4 CATGCCATCCCATCGACGAGTCCATCGTGAATCCCGGCTCTCCGCTTCCCTGACG 63  
 Qy 16 SerValAlaValThrGlnGlnLeuTyrAlaGlnProAenGluSerLeuProThrValGlu 35  
 Db 64 GCGCTCGCGCGCACCTTCCCTTGACACACCGCCATCGACAGCATCC---CGCGAA 120  
 Qy 36 LeuGluProValValIleThrIleAspLysSerGlyMetAlaLeuAlaAenArgIleThr 55  
 Db 121 CTCGATACCGTGGTGTCTACCGCCACCGCAC-----GAAACCCGAGCGCG 168  
 Qy 56 GlnMetProHisThrThrLysValIleTyrGluGluGlnIleGlnGluAlaThrGly 75  
 Db 169 CAGTTCGCGCAGTCCGTGACCGCTCTGACCGCGGAGGACATCAAGGAGCGGGGTCTAT 228  
 Qy 76 SerArgGlnLeuAlaAepValMerAlaGlnLeuIleProSerLeuGlyValSer---Ser 94  
 Db 229 TCG-----GTCGACACAGATCTCGGCATCTCGTTCGCGGTCTCGACATGCCAGAGC 279  
 Qy 95 Gly-----ThrThrSerAenPheGlyGlnThrMet 104  
 Db 280 GCGCGTCCGCGACGGGAAACCTCGTGTCTCGCGCGCGCAACTCCGCGCAGACCATC 339  
 Qy 105 HisGlyArgGlnValGlnPheLeuAenGlyValProLeuThrGlySerArgAspIle 124  
 Db 340 -----GTCCTGATCGACTGGACCGA-ATGACACGCCCTCTCGCGCAACGCGGGTT 392  
 Qy 125 SerArgGlnLeuAenSerIleAenProAenGln-ValAlaArgIleGluValLeuSerGl 144  
 Db 393 -----CGACTTCGCAACTTGACCATCGACACATCGACGCGCATCGAAATCTCGCGCG 446  
 Qy 144 Y---AlaThrSerIleTyrGlySerGlyAlaThrGlyLeuIleAenIleValThrLys 163  
 Db 447 CCGCGGAGCTCGATCTATGTTCCGAAGCCATCGCGGAGTAAATCAACATCATCACCA 506  
 Qy 163 sSerAspLeuGluGluGlnPheGluThrArgIleGlyValHisGlySerLysLeuSe 183  
 Db 507 GNAGGGCTCTGGCAACCGAGGCTGAAACCC-----AGCGTCGACGCGCGGTAGTAGCA 560  
 Qy 183 sSerGluGlyIleGlyTyrGlnValGlyGlnSerValAlaGlyValSerGluAenGlyAs 203  
 Db 561 CACT-----TGGCGAGTCATGGGCGACCTCGCGCGCGGATCCCGCTTCAA 608  
 Qy 203 nValLeuAlaArgLeuAepValAepTyrArgThrThrGlyAlaPheAepAlaAenGl 223  
 Db 609 C-----TACAGCTCACCCCGCATCCCGCGAGTCCCGCGG 644  
 Qy 223 YLysArgIleAlaProGluProAlaGlnThrAspLysGlnAepSer---LysSerLeuSe 242  
 Db 645 CTTTTCGCGCGCGACGACACTTTCGGCGCATCTCGCGCGCGGTTACCGCAACACCA 704

QY 242 rValaantThrAsnValAspTrpGlnLeuAspLysGlnAsnIleAsnLeuAlaLeuTh 262  
 DB 705 CGTACACACCGGATCGCGGACACCGCGGACGAGCTGAGTGGGCTGAGCGCTG-- 762  
 QY 262 rHisTyrAsnAsp-----LysGlnAspThr--AspTyrAla----- 273  
 DB 763 -CGCTACAGAGCGGCTTCAACCAAGTGAATTAAGACTCTGCTCGCGGACCGCT 821  
 QY 274 -----ProAspTyrGlyAsnArgLeuAlaValLeuPheGlyGluLysProSerLeuAs 291  
 DB 822 GATGACCCCACTATACCGGCAACCGCAACGAGCTCTAT----- 861  
 QY 291 nAlaIleLysGlyLeuSerLeuSerLugInProLysThrLysSerThrPheAsnI 311  
 DB 862 ----ACCGCGGCTTCGCGC-----ACCCTCAAGCTCTTC----- 891  
 QY 311 eAenTyrHisHisAspAspLeuTrpGlyAsnThrIleAsnThrAsnAlaTyrTyrArg 331  
 DB 892 -----GACAGCGCTGTGGAGCAGCGGTGGAAATG--GCCCTACACCGCGTGT 935  
 QY 331 gGluLys-----GlyArgPheTyrProPheValAlaProph 343  
 DB 936 GAGCGCGCAGTACGACATCGCGGCAACCGGCGATCCCTCCCATTCGCGCGCGCTA 995  
 QY 343 e 343  
 DB 996 C 996  
 RESULT 36  
 AA231949  
 ID AA231949 standard; DNA; 8266 BP.  
 XX  
 AC AA231949;  
 XX  
 DT 26-JAN-2000 (first entry)  
 XX  
 DE M. catarrhalis strain 4223 tbpA-ORF3-tbpB gene locus.  
 XX  
 KM TbpB gene; Tbp2; transferrin binding protein; diagnosis; otitis media;  
 XX genetic immunization; Moraxella infection; antigen; vaccine; detection;  
 KM antitumor antibody production; therapy; ss.  
 XX  
 OS Moraxella catarrhalis.  
 XX  
 PN MO9952947-A2.  
 XX  
 PD 21-OCT-1999.  
 XX  
 PF 12-APR-1999; 99MO-CA00307.  
 XX  
 PR 14-APR-1998; 98US-0059584.  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 XX  
 PI Myers IE; Schryvers AB; Harkness RE; Loomore SM; Du R; Yang Y;  
 XX Klein ME;  
 DR WPI; 1999-620376/53.  
 DR P-PSDB; AAY43381, AAY43382, AAY43383.  
 PT Nucleic acid encoding transferrin binding protein 2 of Moraxella  
 CC catarrhalis, useful for diagnostics, immunization and recombinant  
 CC protein production -  
 CC  
 PS Example 6; Fig 8; 114pp; English.  
 XX  
 CC This sequence represents the Moraxella catarrhalis strain 4223  
 CC tbpA-ORF3-tbpB locus, and encodes the LPS1 transferrin binding protein  
 CC (Tbp2) of the invention. This sequence is also referred to as the TbpB  
 CC gene. The TbpB gene is used to produce recombinant Tbp2;  
 CC for identification or diagnosis of Moraxella, or for cloning related  
 CC species, using hybridisation assays; and for genetic immunisation against

CC Moraxella infections, e.g. otitis media. The Tbp2 proteins are useful as  
 CC antigens, either in vaccines (including components of conjugate vaccines  
 CC that contain antigens from other bacteria or from tumors, in which case  
 CC they elicit production of antitumor antibodies that may be coupled to  
 CC chemotherapeutic agents or biologically active agents) or to raise  
 CC antibodies (for use as diagnostic reagents and for treating Moraxella  
 CC infections), also for detecting Moraxella antibodies.  
 XX  
 SQ Sequence 8266 BP; 2754 A; 1754 C; 1683 G; 2075 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 0.0122 Length: 8266  
 Score: 138.00 Matches: 95  
 Percent Similarity: 35.57% Conservative: 48  
 Best Local Similarity: 23.63% Mismatches: 149  
 Query Match: 6.74% Indels: 110  
 DB: 20 Gaps: 21  
 US-09-889-746-2 (1-400) x AA231949 (1-8266)  
 QY 10 LeuSerLeuProLeuLeuSerValAla-----ValThrGlnGlnLeu 23  
 DB 133 TTTCTTTGGGTCTGCTTAATCAATCAGCGAGTGGCAGTGGCAACACGCGGATAG 192  
 QY 24 TyrAlaGlnProAsnGlnSerLeuProThrValGluLeuGlnProValValIleThrIle 43  
 DB 193 GCGGAGGCAACAGATAAGCAAACTTGTCTTGTCTTGATGATAAATCTGTGAACAGCG 252  
 QY 44 AspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrThrLysVal 63  
 DB 253 AAGAAAAACGCCCGTAAA--GCCAAGCAAGATTACGCGCTGTGAAGGTGTCAAAACT 309  
 QY 64 IleTyrGluGlnIleGlnGlnIleAlaThrGlySerArgGlnLeuAlaAspValMet 83  
 DB 310 GCC--GAGACCATCAATTAAGAACAGTCTAAACATTGAGACTTA----- 354  
 QY 84 AlaGlnLeuIleProSerLeuGly--ValSerSerGlyThrThrSerAsnPheGln 102  
 DB 355 ACACGCTATGACCCCTGGCATTCGTGTGTGAAGCAAGTGTGGGCAAGCTGAGCTAT 414  
 QY 103 ThrMetHisGly-----ArgGlnValGlnPheLeuLeuAsnGlyVal----- 116  
 DB 415 TCTATTCGTGTGTATGATAAATCGTGTGGCGTATGCTTATGATGATCAATCAAGCC 474  
 QY 117 -----ProLeuThrGlySerArg--AspIleSerArgGlnLeu 128  
 DB 475 CAGCACTATGCCCTTACAGAGCCCTGTGGCAGCAAAATTAATGCCAGGTGGGCAATC 534  
 QY 129 AsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyAla---ThrSer 147  
 DB 535 AAGCAATTAAGAAATACGAAATATGCCCTCCGTTGAGATTAAGTAAAGTCCAAATTCAAGT 594  
 QY 148 IleTyrGlySerGlyValaThrGlyGlyLeuIleAsnIleValaThrLysSer----- 164  
 DB 595 GAATACGGCTCTGGGCATTAATCTGCTGTGGCATTTGTATCAAAACCGCGATGAC 654  
 QY 165 -----AspLeuGlnGlnGlnGlnPheGlnThrArg----- 174  
 DB 655 ATCATCAAGATGTAAAGATTGGGGCGTGCAGACCAAAACCGCTATGCCACTAAAT 714  
 QY 175 -----IleGlyValHisGlySerLysLeuSerSerGlnGlyIle--- 187  
 DB 715 AAGCATGGTTAATTTCTGTGTGGAGAGCAGCAGCAAGCAGGTTCTTTAGCGGCTTATC 774  
 QY 188 -----GlyTyrGlnValGly 192  
 DB 775 ATCTACACGACCGCGTGTCAAGATACAAAGCAGATGATGATCTATCAGGGTAGC 834  
 QY 193 GlnSer-----ValAla 196  
 DB 835 CAAGTTTGTATAGACCGGTGGCAACCACTGACCAATTAACGAACTTTTAATAGCA 894  
 QY 197 GlyValSerGlyAsnGlyAsnVal-----LeuAla 206

```

Db 895 AATGAATGTCACATGGTAAATTATGAGCGGTGTGCTGCGCGGTCAAAACCAACTTCAA 954
Qy 207 ArgLeuAspValAspTyrArgThrThrGlyGlyAlaPheAspAlaAsnGly---LysArg 225
Db 955 GCCAAGCCCAACCAATGGTGGTGAAGTCATGTCAGATGTCAGAGTCTTACACCGC 1014
Qy 226 IleAlaProGluProAlaGlnThrAspLysGlnAspSerLysSerLysSerValAsnThr 245
Db 1015 CTTATCCCAACCCACTC-----ACCCAAGACAGCAAACTCTTACTGCTTCCGCCA 1065
Qy 246 AsnValAspTrpGlnLeuAspLysGlnAsnIleAsn-----LeuAlaLeuThrHis 263
Db 1066 GGT-----TATCAGCTAAACGATAGCACTATGTCGGTGTGTATGAATCACCAAA 1119
Qy 264 TyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeu 283
Db 1120 CAAAACCTAGCCATCAAGATAAAACCGTCTCTTAT-----CTGGCGGTTTCAT 1170
Qy 284 PheGlyGluLysProSerLeuAsn-----AlaIleLysGlyLeu 296
Db 1171 GACATTGAAAATCAAGGCTCAGCAACCATGCCAAGCAATGGCTATTATCAAGGCAAT 1230
Qy 297 SerLeuSerGluGlnProLysThrThr-----LysSerThrPheAsnIleAsnTyr 313
Db 1231 AATCTTGTGTAACGCAATTCGTGATACCATTTGGCCAGATTCAGGTTATGGCATCAACTAT 1290
Qy 314 HisHis 315
Db 1291 GCTCAT 1296
RESULT 37
AAT10105
ID AAT10105 standard; DNA; 8041 BP.
XX AAT10105;
AC AAT10105;
XX
DT 13-MAY-1996 (first entry)
XX
DE Adherence conferring plasmid pear.
XX
KW adhesin; plasmid pear; vector; vaccine; intestine colonisation; ds.
XX
OS Chimeric Escherichia coli;
OS Chimeric 'synthetic.
XX
FH Key Location/Qualifiers
FT primer_bind complement (2867..2890)
FT primer_bind /tag= a
FT primer_bind /note= "primer for E.coli O157:H7 DNA"
FT primer_bind complement (2895..2914)
FT primer_bind /tag= b
FT primer_bind /note= "primer for pear"
FT CDS 3036..5126
FT /tag= c
FT /product= adhesin
FT /trans_except_pos=3912..3914; aa:Gly
FT /note= "claim 1, page 36"
FT misc_feature 3271..3310
FT /tag= d
FT /note= "region of transposon TnpHoA insertion"
FT misc_feature 3801..3840
FT /tag= e
FT /note= "region of transposon TnpHoA insertion"
FT primer_bind complement (5176..5196)
FT /tag= f
FT /note= "primer for pear"
FT primer_bind complement (5159..5183)
FT /tag= g
FT /note= "primer for E.coli O157:H7 DNA"
FT CDS complement (6449..7024)
FT /tag= h
FT /note= "terE homologue"

```

```

CDS complement (7092..7670)
FT /*tag= i
FT /note= "terD homologue"
PN W09600233-A1.
XX
PD 04-JAN-1996.
XX
PP 07-JUN-1995; 95WO-US06994.
XX
PR 24-JUN-1994; 94US-0265714.
XX
XX (CHIL-) CHILDREN'S HOSPITAL & MEDICAL CENT.
XX (UNIW ) UNIV WASHINGTON.
XX (UNIW ) UNIV WASHINGTON STATE RES FOUND.
XX
XX Besser TE, Bilge SS, Tarr PI, Vary JC;
XX
XX WPI; 1996-068826/07.
XX P-PSDB; AAR75366.
XX
XX Chromosomal DNA from E. coli O157:H7 encoding epithelial adhesin -
XX isolated on plasmid pSC (overlap), for use as a vaccine to prevent
XX bacterial colonisation of bovine intestine
XX
XX Disclosure; Page 23-30; 42pp; English.
XX
XX Adherence conferring plasmid pear (AAT10105) comprises Escherichia
XX coli O157:H7 chromosomal DNA plus the Stratagene SK+ vector.
XX It includes 3 open reading frames, 2 of which are homologues of
XX terE and terD genes necessary for cellulite resistance. The
XX third ORF is homologous to the IrgA gene and encodes an adhesin
XX (AAR75366) that enables E. coli O157:H7, an antibiotic-resistant,
XX virulent and common food-borne pathogen, to adhere to epithelial
XX cells. This ORF can be utilised in the prodn. of adhesin for use
XX as a vaccine to prevent disease or colonisation of mucosal surfaces
XX by O157:H7.
XX
XX Sequence 8041 BP; 2241 A; 1807 C; 1838 G; 2126 T; 29 other;
XX
Alignment Scores:
Pred. No.: 0.0183 Length: 8041
Score: 136.00 Matches: 74
Percent Similarity: 37.46% Conservative: 53
Best Local Similarity: 21.83% Mismatches: 118
Query Match: 6.64% Indels: 94
DB: 17 Gaps: 18
XX
US-09-889-746-2 (1-400) x AAT10105 (1-8041)
Qy 40 VallleThrIleAspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHis 59
Db 3114 GTGATGATTGCTCGGCATCCGGC-----TATGAGAAAAAGCTGACTAACGCAGCCGCC 3167
Qy 60 ThrThrLysValIleTyrGluGluGlnIleGlnGluAlaThrGlySerArgGlnLeu 79
Db 3168 AGTGTTTCTGTGATTAGCCAGGAGGAATTGCAG-----TCCAGCCAGTAC 3212
Qy 80 AlaAspValMetAlaGlnLeu-----IleProSerLeuGlyValSerSerGlyThrThr 97
Db 3213 CACGATCTGGCGGAGGCTCTGAGATCAGTAGAGGTGTGGATGTGAAAGTGGTACGGGT 3272
Qy 98 SerAsnPheGly-----GlnThrMetHisGlyArgGlnValGlnPheLeu 112
Db 3273 AAAACCGGAGGCTGGAATCAGCATCCGAGGAATCCAGCAGCAGTTACACGCTGATCTG 3332
Qy 113 LeuAsnGlyValProLeuThrGlySerArgAspIleSer---ArgGlnLeuAsnSerIle 131
Db 3333 ATTGATGTGTTCGTTCAGGGCGGGAAGCAGTACGTCACCGGTTTCTGCCCATG 3392
Qy 132 AsnProAsn-----GlnValAlaArgIleGluValLeuSerGlyAla 145
Db 3393 AATACCGGGTTCATGCCCTCTCTGCCGCCCATTTGAGCGTTATCAGGGGGCGG 3452

```

```

QY 146 ThrSer---IleTyrGlySerGlyAlaThrGlyLeuIleAsnIleValThrLysSer 164
DB 3453 ATGTCACACCTGATGCTCTGATGGATGGGGGTGTGTGAATATCATTCACAGAAAG 3512
QY 165 AspLeuGluGluGluGlnPheGluThrArgIleGlyValHis----- 178
DB 3513 AATGACAGACAAATGGCTCTCTCCGTAATGACGGCTGTGACGAAACAAACAA 3572
QY 179 -----GlySerLysSerSerGluGlyIle 187
DB 3573 TGGGGTAACGACGCCACTTTAATTTCGAGACAGTGGTCCCTGTGTGATGATTCGTG 3632
QY 188 GlyTyrGlnVal-----GlyGlnSerValAlaGlyValSer 199
DB 3633 AGCTTCAGGTACGGGTAGACACAAACGCGTACGGTTCATCGGTACATCAGTACGAGC 3692
QY 200 GluAsnGlyAsnValLeuAlaArgLeuAspValAspTyrArgThrGlyAlaAlaPhe 219
DB 3692 ----- 3692
QY 220 AspAlaAsnGlyLysArgIleAlaProGluProAlaGlnThrAspLysGlnAspSerLys 239
DB 3693 GATACACGACGGCAGCGGTATTT---CCTTATCCACG-----GAGTCACAG 3734
QY 240 SerLysSerValAsnThrAsnValAspTyrGlnLeuAspLysGlnAsnIleAsnLeu 259
DB 3735 AATTATATATCTTGCTGACACTCTTGACTGSAAGCGTCGAGACAGATGCTC----- 3788
QY 260 AlaLeuThrHisTyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArg 279
DB 3789 -----TGCTTTGATGATGATGATACACG-----CGGACAGCGTATGATTAACCGG 3830
QY 280 LeuAlaValLeuPheGlyGluLysProSerLeuAsnAlaIleLysGlyLysSerLysSer 299
DB 3831 -----GATGGGCAACTGGGGAGTCTGACGGGGGATATGACCGACCCCTG--- 3875
QY 300 GluGlnProLysThrThrLysSerThrPheAsnIleAsnTyrHisAsp----- 316
DB 3876 -----CGCTATGACGCAAAACAAATTTCAGCTGCTATGATCATCTTCACCTTC 3926
QY 317 AspLeuTyrGlyAsnThrIleAsnThrAsnAlaTyrTyrArgArgGluLysGlyArg 335
DB 3927 GGAACATGGAATCTGTATCTGAACGACGAG-----ACAGAAATTAAGTCTGT 3977

RESULT 38
ABA88835
ID ABA88835 standard; DNA; 2139 BP.
AC ABA88835;
XX
XX
DT 11-FEB-2002 (first entry)
DE Escherichia coli polynucleotide SEQ ID NO 262.
XX
XX
XX Escherichia coli; B2/D+A-; anti-inflammatory; antibacterial;
XX immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
XX systemic infection; non-diarrhoeal infection; septicemia;
XX pyelonephritis; antibiotic resistance; ds.
XX
XX Escherichia coli.
XX
XX MO200166572-A2.
XX
XX 13-SEP-2001.
XX
XX 12-MAR-2001; 2001MO-EP03445.
XX
XX 10-MAR-2000; 2000PR-0003145.
XX 02-FEB-2001; 2001PR-0001449.
XX
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX

```

```

PI Bingen E, Bonaccorsi S, Clermont O, Nassif X, Tinsley C;
XX WPI; 2001-550253/61.
XX
XX A library of DNA fragments of Escherichia coli strains for the
PT phylogenetic determination of a given strain comprises polynucleotides of
PT nature B2/D+ A- -
XX
XX Example 6; Fig 6; 646bp; English.
XX
XX The invention relates to a library of DNA fragments of Escherichia coli
CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533) of nature
CC and encoded proteins (ABBS2459-ABBS2919 and ABBS2954-ABBS3094) of nature
CC B2/D+A-. The polynucleotides have potential anti-inflammatory,
CC antibacterial and immunosuppressive activity as part of pharmaceutical
CC compositions used to treat, palliate or prevent extra-intestinal E. coli
CC infections. The polypeptides are useful for determining the phylogenic
CC group of a given E. coli strain. These polypeptides can detect and treat
CC an undesired development of E. coli, particularly an extra-intestinal
CC infection that include systemic and non-diarrhoeal infections such as
CC septicemia, pyelonephritis and meningitis this is particularly
CC advantageous as bacterial resistance is increasing with the more
CC frequent use of broad spectrum antibiotics.
XX
XX Sequence 2139 BP; 584 A; 519 C; 508 G; 528 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 0.0038 Length: 2139
Score: 135.50 Matches: 87
Percent Similarity: 36.77% Conservative: 45
Best Local Similarity: 24.23% Mismatches: 132
Query Match: 6.62% Indels: 95
DB: Gaps: 22

US-09-889-746-2 (1-400) x ABA88835 (1-2139)
QY 18 AlaValThrGlnGlnLeuTyrAlaGlnPro-----AsnGluSerLeuProThr 33
DB 43 GGGCTTTTCACTGCTTTTGGGCGACCGATGATTCAGCAACAGCTCTGTAACGACC 102
QY 34 ValGluLeuGluProValAlaIleThrIleAspLysSerGlyMetAlaLeuAlaAsnArg 53
DB 103 AAGATGGCGCAACATCACTGCTTACGACGAT-----CCAAATACC 144
QY 54 IleThrGlnMetProHisThrThrLysValIleTyrGluGluGlnIleGlnGluAla 73
DB 145 GCAACTGAGGCA-----ACCGATGTTATTAACCTCTGAC 180
QY 74 ThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSer----- 89
DB 181 ACCTCCACCGGACATTAACCGATATGCGGATGCGATATCCGACGTCGTCATACG 240
QY 90 -----LeuGlyValSerSerGlyThrThr----- 97
DB 241 GTTAGGATCAGGTTCTGGAACAAACAGAAATGCGACAAACCTGATGAGCGCTTTATAC 300
QY 98 -----SerAsnPheGlyGln-----ThreHisGlyArgGlnValGlnPheLeuAsn 114
DB 301 GTTCGTAACGGTGTACGACCAATATACATTAAGCGGAGCTCAGAGTCTTTGTACCGCT 360
QY 115 GlyValProLeuThrGlySerArgAsp-----IleSerArgGlnLeuAsnSerIle 131
DB 361 GGG-----TTTGGCGCAACCGGATGGCTCCATCATATGACCAACGCTCTGCAACGTA 414
QY 132 AsnProAsnGln-----ValAlaArgIleGlnValLeuSerGly---AlaThr 146
DB 415 CTTCCTCTGATGTTCAACCGCGCAACAGACGAGCGTGTGAAGTGTAAAGCGCGCTCC 474
QY 147 SerIleTyrGlySerGlyAlaThrGlyLysLeuIleAsnIleValAlaThrLysSerAspLeu 166
DB 475 ACCTGTATGGCACTTCGATCTCGGAGCTGATTAACGTCGTGACCAAGCCC----- 528
QY 167 GluGluGluGlnPheGluThrArgIleGlyValHisGlySer---LysLeuSerSerGln 185

```

```

Db 529 CCGGAAAAACATTC-----CATGGTTCGGTTCCAGCCACCTCCCTCC 570
Qy 186 GlyIleGlyTyrGlnValGlyGln---SerValAlaGlyValSerGluAsnGlyAsnVal 204
Db 571 AGTTTGTGGCGGCACATGGGCACTTGTATCATCAGAGTCCCATTTGAA---GGCACTCAG 627
Qy 205 LeuAlaArgLeuAspValAspTyrArgThrThrGlyGlyAlaPheAspAlaAsn----- 222
Db 628 CTGGCG-----TATCGCTTACCGGGAAGTCCAGATCAAGATTACTGG 672
Qy 223 -----GlyIysArgIleAlaProGluProAlaGlnThrAspLysGlnAspSerLys 239
Db 673 CGAAACTTCGGTAAA-----GAGCGC 693
Qy 240 SerLeuSerValAsnThrAsnValAspTyrGlnLeuAspAspLysGlnAsnIleAsnLeu 259
Db 694 AGTACATTATTGCCCCGCTCACTCACTGG---TTTGGTGAATGAACACAGTAACCATG 750
Qy 260 AlaLeuThrHisTyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArg 279
Db 751 CTCATTATCCCAI-----CGGACTATAAACTCCATTCGATCGTGGA 792
Qy 280 LeuAlaValLeuPheGlyGlyLysProSerLeuAsnAlaIleLysGlyLeuSerLeuSer 299
Db 793 ACGATTTTCGACCTTACGACGAAACAGCCCGTAAACGTTGATCGAAAAATACGTTTTCAC 852
Qy 300 GluGlnProLysThrThrLys-----SerThrPheAsnIleAsnTyrHis 314
Db 853 GAACCGTTTAAATTACAGATCGTCAGTCGATCTGGCGCAACTCAACGCAGAAATATCAT 912
Qy 315 HisAspAspLeuTyrGlyAsnThrIleAsnThrAlaAsnAlaTyrTyrArgArgGlyLys 333
Db 913 CTCATAGCAGCTGG-----ACAGCGCGCTTGTATTACAGCTACAGCAGGATATAA 963

RESULT 39
ID ABA88834 standard; DNA; 7315 BP.
XX AC ABA88834;
XX DT 11-FEB-2002 (first entry)
XX DE Escherichia coli polynucleotide SEQ ID NO 261.
XX KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
XX KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
XX KW systemic infection; non-diarrhoeal infection; septicemia;
XX KW pyelonephritis; antibiotic resistance; ds.
XX OS Escherichia coli.
XX PN W0200166572-A2.
XX PD 13-SEP-2001.
XX PF 12-MAR-2001; 2001WO-EP03445.
XX PR 10-MAR-2000; 2000PR-0003145.
XX PR 02-FEB-2001; 2001PR-0001449.
XX XX
XX FA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX XX Bingen E, Bonacorei S, Clermont O, Nassif X, Tinsley C;
XX XX WPI; 2001-550253/61.
XX DR
XX PT A library of DNA fragments of Escherichia coli strains for the
XX PT phylogenetic determination of a given strain comprises polynucleotides of
XX PT nature B2/D+ A-
XX XX
XX PS Example 6; Fig 6; 646pp; English.
XX XX

```

```

CC The invention relates to a library of DNA fragments of Escherichia coli
CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
CC and encoded proteins (ABBS2459-ABBS2919 and ABBS2954-ABBS3094) of nature
CC B2/D+A-. The polynucleotides have potential antiinflammatory,
CC antibacterial and immunosuppressive activity as part of pharmaceutical
CC compositions used to treat, palliate or prevent extra-intestinal E. coli
CC infections. The polypeptides are useful for determining the phylogenetic
CC group of a given E. coli strain. These polypeptides can detect and treat
CC an undesired development of E. coli, particularly an extra-intestinal
CC infection that include systemic and non-diarrhoeal infections such as
CC septicemia, pyelonephritis and meningitis this is particularly
CC advantageous as bacterial resistance is increasing with the more
CC frequent use of broad spectrum antibiotics.
XX
XX SQ Sequence 7315 BP; 2094 A; 1928 C; 1657 G; 1636 T; 0 other;

Alignment Scores:
Pred. No.: 0 018 Length: 7315
Score: 135.50 Matches: 87
Percent Similarity: 36.77% Conservative: 45
Best Local Similarity: 24.23% Mismatches: 132
Query Match: 6.62% Indels: 95
DB: 22 Gaps: 22

US-09-889-746-2 (1-400) x ABA88834 (1-7315)
Qy 18 AlaValThrGlnGlnLeuTyrAlaGlnPro-----AsnGluSerLeuProThr 33
Db 5149 GCGCTCTTTTCACTGCTTTTGGCGCACCGATGATTCATGCAACAGACTCTGTACGACC 5208
Qy 34 ValGluLeuGluProValValIleThrIleAspLysSerGlyMetAlaLeuAlaAsnArg 53
Db 5209 AAAGATGCGGAACAACATCACTGTTACAGCAGAT-----GCAATATACC 5250
Qy 54 IleThrGlnMetProHisThrThrIleValIleTyrGluGlnIleGlnGluAla 73
Db 5251 GCAACTGAGGCA-----ACCGATGGTTATCAACCTCTGAGC 5286
Qy 74 ThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSer----- 89
Db 5287 ACCTCCAGCGGACATTAACCGATATGCGGATATCCGCGAGTGTGTCATACG 5346
Qy 90 -----LeuGlyValSerSerGlyThrThr----- 97
Db 5347 GTTAGCGATCAGGTTCTGGAAACCCAGAAATCGCAACAGCTGGATGAGCGCTTTATAAC 5406
Qy 98 ---SerAsnPheGlyGln-----ThrMethIleGlyArgGlnValGlnPheLeuLeuAsn 114
Db 5407 GTCAGTAACGTGTACAGCAACCAATACATTAGCGGAGCTCAGGATGCTTTTGTACCGCGT 5466
Qy 115 GlyValProLeuThrGlySerArgAsp-----IleSerArgGlnLeuAsnSerIle 131
Db 5467 GGG-----TTTGGCGCAACCGGATGCTCCATCATGACCAACGGTCTCGGAACCGTA 5520
Qy 132 AsnProAsnGln-----ValAlaArgIleGluValLeuSerGly---AlaThr 146
Db 5521 CTTCTCTGCTAGTTTCAACGCGCAACAGAGCGTGTGGAAGTGTCTGAAAGGCGCGCTCC 5580
Qy 147 SerIleTyrGlySerGlyAlaThrGlyLeuIleAsnIleValThrLysSerAspLeu 166
Db 5581 ACGCTGTATGGCATTTCTGATCCTCGGCGACTGATTACGTCGTGACCAAGCGC----- 5634
Qy 167 GluGluGluGlnPheGluThrArgIleGlyValHisGlyIysSer---LysLeuSerSerGlu 185
Db 5635 CCGGAAAAAACATTG-----CATGCTCGGTTTTCAGCCACCTCTCC 5676
Qy 186 GlyIleGlyTyrGlnValGlyGln---SerValAlaGlyValSerGluAsnGlyAsnVal 204
Db 5677 AGTTTGTGGCGGCACCTGGGCAACTTGTATCATCAGAGTCCCATTTGAA---GGCACTCAG 5733
Qy 205 LeuAlaArgLeuAspValAspTyrArgThrThrGlyGlyAlaPheAspAlaAsn----- 222
Db 5734 CTGGCG-----TATCGCTTACCGGGAAGTCCAGATCAAGATTACTGG 5778

```



QY 223 -----GlytysArgilealaProgluProAlaGlnThrAspIysGlnaSerIys 239  
 DB 5779 CGAACTCGGTAA-----GAGCGC 5799  
 QY 240 SerLeuSerValaenThrAsnValaAspTrpGlnLeuAspAspIysGlnaenIleu 239  
 DB 5800 AGTACATTATATCCCGCTACCTCCTGG--TTGGTATATATGACMACAGTAAACCATG 5856  
 QY 260 AlaLeuThrIstIysAsnAspIysGlnaAspThrAspIysAlaProaAspIysArg 279  
 DB 5857 CTCTATTCAT-----CGGACCTATTAACCTCATTCGATCCGGA 5898  
 QY 280 LeuAlaValaLeuPheGlyGlyIysProSerLeuAsnAlaIlePheGlySerLeuSer 239  
 DB 5899 ACATTTTCGACCTTACGACGAAACAGCCGTAAACGTTGATCGAATAATACGTTTGAC 5958  
 QY 300 GluGlnProIysThrIlys-----SerThrPheAsnIleAsnIleAsnIle 314  
 DB 5959 GAACCGTTTAAATATTCAGATGTCAGTCCGATCTGCGCACTCAACGCAATATCAT 6018  
 QY 315 HisAspAspLeuTrpGlyAsnThrIleAsnThrAsnAlaIysTrpArgAspGlyIys 333  
 DB 6019 CTCAATATGACAGTGG-----ACAGCGCGCTTGATTACAGTACAGCCAGATAAA 6069  
 RESULT 40  
 AAS93059  
 ID AAS93059 standard; cDNA; 2727 BP.  
 AC AAS93059;  
 DT 13-FEB-2002 (first entry)  
 DE DNA encoding novel human diagnostic protein #28863.  
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KM food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 OS Homo sapiens.  
 PN WO200175067-A2.  
 PD 11-OCT-2001.  
 PF 30-MAR-2001; 2001WO-US08631.  
 PR 31-MAR-2000; 2000US-0540217.  
 PR 23-AUG-2000; 2000US-0649167.  
 PA (HYSB-) HYSBQ INC.  
 PI Drmanac RT, Liu C, Tang YT;  
 DR P-PSDB; ABG28872.  
 PT New isolated polypeptide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 PS Claim 1; SEQ ID No 28863; 103bp; English.  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences, (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostic, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pat\_sequences.  
 XX  
 SQ Sequence 2727 BP; 627 A; 677 C; 791 G; 631 T; 1 other;  
 Alignment Scores:  
 Pred. No.: 0.00526 Length: 2727  
 Score: 135.00 Matches: 63  
 Percent Similarity: 40.62% Conservative: 54  
 Best Local Similarity: 21.88% Mismatches: 105  
 Query Match: 6.60% Indels: 66  
 DB: 23 Gaps: 14  
 US-09-889-746-2 (1-400) x AAS93059 (1-2727)  
 QY 14 LeuLeuSerValaAlaValaThrGlnGlnLeuIleTyrAlaGlnProaGlnLeuSerProThr 33  
 DB 856 TTGTCATCTGGGATTTATGGGCTAGCGGACGACAAAGCCGATACCTCTGTT 915  
 QY 34 ValGlnLeuGlnProValaIleThrIleAspIysSerGlyMetAlaLeuAlaAsnArg 53  
 DB 916 TCACATGACGATATCTATTTGCTTACC-----GCCGCCGACGAG 954  
 QY 54 IleThrGlnMetProIleThrIleThrIleValIleTyrGlnGlnIleGlnGlnAla 73  
 DB 955 AACTTACAGCGCCCTCGCTTGCACCATCACCAGATGAA--ATCCGAAAAACCCG 1011  
 QY 74 ThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSer 93  
 DB 1012 GTT---GCCCGCATGTGTGCAAGATCATC---CGTACATGCGACGCTTACCTGACC 1065  
 QY 94 SerGlyThrThrSerAsnPheGlyGlnThrMetHisGlyArgGlnValGln----- 110  
 DB 1066 GGTACTCCACAGT-----GGTCAGCGTGGGAATTAACGACAGATGATATTCGCGGT 1119  
 QY 111 -----PheLeuLeuAsnGlyValProLeuThr----- 119  
 DB 1120 ATGGCTCCGAAAAACGCTGATTGTGATTGACGCGAAGCCGGTAAAGCCGTAATCTG 1179  
 QY 120 -----GlySerArgAspIleSerArgGlnLeuAsnSerIleAsnPro 133  
 DB 1180 GTGCGTCAGGGCTGGCGTGGCGAGCCGATACCCGCTGTGATCTTCTGGGCGCACCT 1239  
 QY 134 AsnGlnValAlaArgIleGlnValLeuSerGly--AlaThrSerIleTyrGlySerGly 152  
 DB 1240 GAATGATTGAACGTATTGAAGTTCTGCGTGGCCGCAAGTCCGTTATGCAACGCGC 1299  
 QY 153 AlaThrGlyGlyLeuIleAsnIleValThrIysSerAspLeuGlnGlnGlnPheGln 172  
 DB 1300 GCGCGCGCGCGCTGTATTAACATCTTACCAAAAAAGCAGCGCGAG----- 1347  
 QY 173 ThrArgIleGlyValHisGlySer-----LysLeuSerSer 184  
 DB 1348 -----TGGCAGCGCTCTGGAGCGCATATTTCATGCGCCAGACATAAAGAG 1395  
 QY 185 GlyGlyIleGlyTyrGlnValGlyGlnSerValAlaGlyVal-----SerGluAsn 201  
 DB 1396 GAAGGTGCCACCAAAACGACCTAATCTTACCGCGTCCGCTGGCGCAGATTCAGC 1455  
 QY 202 GlyAsnValLeuAlaArgLeuAspValaAspIysArgThrThrGlyGlyAlaPheAspAla 221  
 DB 1456 TTCCGTTGTATGGCAACCTCGCAAAACCCAGGTACCGTGGGTATACACAGGCGC 1515  
 QY 222 AsnGlyLysArgIle-----AlaProGluProAlaGlnThrAspIysGlnaSer 238

Db 1516 ATCAGTCGCGGTCCGGAACGATGACACAGCTTACCGCCGCGGAGGGGTAAAT 1575  
 Qy 239 LysSerLeuSerValAsnThrAsnVal-----AspTyrGlnLeuAspAspLysGlnAsn 256  
 Db 1576 CAATTCTTCATATTCATACATGATCTCGGAAATGACGATA----- 1620  
 Qy 257 IleAsnLeuAlaLeuThrHisTyr 264  
 Db 1621 -----TTTGGCGCATTAC 1632

## RESULT 41

AA588838  
 ID AA588838 standard; cDNA; 2808 BP.

AC AA588838;

XX 13-FEB-2002 (first entry)

DT DNA encoding novel human diagnostic protein #24642.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

OS WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YT;

XX WPT; 2001-639362/73.

DR P-PSDB; ABG24651.

XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity

XX Claim 1; SEQ ID No 24642; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridization probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (II) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AA94564 represent novel human  
 CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 2808 BP; 648 A; 704 C; 812 G; 643 T; 1 other;

Alignment Scores:  
 Pred. No.: 0.00547 Length: 2808  
 Score: 135.00 Matches: 63  
 Percent Similarity: 40.62% Conservative: 54  
 Best Local Similarity: 21.88% Mismatches: 105  
 Query Match: 6.60% Indels: 66  
 DB: 23 Gaps: 14

US-09-889-746-2 (1-400) x AAS88838 (1-2808)

Qy 14 LeuLeuSerValAlaValThrGlnGlnLeuTyrAlaGlnProAsnGluSerLeuProThr 33  
 Db 937 TTGTCAATCTGGGATTTATGGGTAGCCGACGACAGAGCCGACGATCTCTCTT 996  
 Qy 34 ValGluLeuGluProValValIleThrIleAspLysSerGlyMetAlaLeuAlaAsnArg 53  
 Db 997 TCACATGACGATACTATTGTCGTTACC-----GCCGCCGAGCAG 1035  
 Qy 54 IleThrGlnMetProHisThrThrLysValIleTyrGluGluGlnIleGlnGluGlnAla 73  
 Db 1036 AACTTACAGGCGCTGGCGTTTCGACCATCACCGCAGATGAA---ATCCGAAAAACCCG 1092  
 Qy 74 ThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSer 93  
 Db 1093 GTT---GCCCGCGATGTGTGAAGATCATC---CGTACCATGCCAGGCGTTAACCTGACC 1146  
 Qy 94 SerGlyThrThrSerAsnPheGlyGlnThrMetHisGlyArgGlnValGln----- 110  
 Db 1147 GGTAATCCACCACT-----GGTCAGCGTGGGAATACCCACAGATTGATTTCGCGGT 1200  
 Qy 111 -----PheLeuLeuAsnGlyValProLeuThr----- 119  
 Db 1201 ATGGGTCCGGAAAAACACGCTGATTTTGATTGACGCGAAGCGGTAAGCAGCGTAACTCG 1260  
 Qy 120 -----GlySerArgAspIleSerArgGlnLeuAsnSerIleAsnPro 133  
 Db 1261 GTGCGTCAGGCGTGGCGGCGGCGATACCCGTGGTGATCTTCTCTGGGTGCCACT 1320  
 Qy 134 AsnGlnValAlaArgIleGluValLeuSerGly---AlaThrSerIleTyrGlySerGly 152  
 Db 1321 GAAATGATTGAACGATTTGAAGTTCTGCGTGTCCGCGNAGTCGCGTTATGGCAACGCGC 1380  
 Qy 153 AlaThrGlyGlyLeuIleAsnIleValThrLysSerAspLeuGluGluGlnPheGlu 172  
 Db 1381 GCGGCGGCGCGTGGTTAATCATATTACCAAAAAAGCAGCGCGAG----- 1428  
 Qy 173 ThrArgIleGlyValHisGlySer-----LysLeuSerSer 184  
 Db 1429 -----TGGCAGCGCTCTGGGACGCGATATTTCATGTGCCGACAGATATAAGAG 1476  
 Qy 185 GluGlyIleGlyTyrGlnValGlyGlnSerValAlaGlyVal-----SerGluAsn 201  
 Db 1477 GAAGTGGCCACCAACGCACTAATTTAGCTGACCGGTCCGCTGGGCGAGCAATTACG 1536  
 Qy 202 GlyAsnValLeuAlaArgLeuAspValAspTyrArgThrThrGlyValAlaPheAspAla 221  
 Db 1537 TTCCGTTTGTATGGCAACCTCGACAAAAACCCAGGTGACGCGTGGGATATCAACACGGGCC 1596  
 Qy 222 AsnGlyLysArgIle-----AlaProGluProAlaGlnThrAspLysGlnAspSer 238  
 Db 1597 ATCAGTCGCGCGTCCCGGACGCGTATGCCAGCGTTTACCAGCGCGCGCGAAGGGGTAAAT 1656  
 Qy 239 LysSerLeuSerValAsnThrAsnVal-----AspTyrGlnLeuAspAspLysGlnAsn 256  
 Db 1657 CAATTCTTCATATTCATACATAAGATCTCTCGMAAATGACGATA----- 1701  
 Qy 257 IleAsnLeuAlaLeuThrHisTyr 264  
 Db 1702 -----TTTGGCGATTAC 1713

## RESULT 42

AA551469

ID AA551469 standard; DNA; 2448 BP.

AC AAS51469;  
XX  
XX 13-FEB-2002 (first entry)  
DT  
DE Pseudomonas aeruginosa DNA for cellular proliferation protein #54.  
DE  
KW Antisense; ds; prokaryotic cellular proliferation gene;  
KM antibiotic; antibacterial; drug design.  
XX  
OS Pseudomonas aeruginosa.  
XX  
PN WO200170955-A2.  
FN  
XX  
PD 27-SEP-2001.  
PF  
XX 21-MAR-2001; 2001WO-US09180.  
XX  
PR 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
PX  
PA (ELIT-) ELITRA PHARM INC.  
PI Haselbeck R, Ohlsen KL, Zykkind JW, Wall D, Trawick JD, Carr GJ;  
PI Yamamoto RT, Xu HH;  
PI  
DR MPI; 2001-611495/70.  
DR P-PsDB; AAU33610.  
XX  
PT New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
PS  
XX Claim 27; Seq ID No 4051; SIlpp; English.

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence encodes an essential prokaryotic cellular proliferation protein.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_gct\_sequences.

Sequence 2448 BP; 557 A; 821 C; 668 G; 402 T; 0 other;

Alignment Scores:

Pred. No.:	Length:
Score:	2448
Percent Similarity:	134.00
Best Local Similarity:	38.75%
Query Match:	21.68%
DB:	6.55%
	Gaps: 16

US-09-889-746-2 (1-400) X AAS51469 (1-2448)

33 ThnvaIGluLeucIubProValIleThrlIeaApIysSergly----- 47

Db 370 AGCGTGAATCTCGGGCGCAACATGATCACTTCCAAACGATTGGGACATCACCGAGGAC 429  
 Qy 48 -----MetAlaLeuAlaAsnArg1LeThr----- 55  
 Db 430 AGCGGCTCTATACGCCGACGACATCTCCGACAGCCCGCTGTGCTCATCCGCGC 489  
 Qy 56 GlnMetProHisThrThrThrLysValIleTyrGlnGlnGlnIleGlnGlnAlaIleThrGly 75  
 Db 490 GAGACGCCCAAGTCATATCACCGTGGTACCCCGCCGAAACATGAGACGAC-----TTGGGC 543  
 Qy 76 SerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSerSerGly 95  
 Db 544 CTCACAAACAATCGACGACGTCATGCGCCAT---ACCGCGGAGTACCGCTTCCGCCCTAC 600  
 Qy 96 ThrThiSerAsnPheGlyGlnThrMetHisGlyArgGlnVal--GlnPheLeuAsn 114  
 Db 601 GACACTACACCGCAACAACATCTATCTGCGCGCGCTTCTCGATCAACAATCTTCAGTACAC 660  
 Qy 115 GlyValProLeuThrGlnGlySerArgAspIleSerArgGlnLeu--AsnSerIleAsnPro 133  
 Db 661 GGGATTCCTCTCGACG---GCCCGCAACGTCGGCTACTCCGCGGCAATACCTCATGTAC 717  
 Qy 134 AsnGlnVal--AlaArgIleGluValIleLeuSerGlyAlaIleThrSerIleTyr---GlySer 151  
 Db 718 ATGGCCATCTACACGCGGGGAGAGTACTCAAGGGGGGACCCGGCTGTCAACCGTCCG 777  
 Qy 152 GlyAlaThrGlyGlyLeuIleAsnIleValThrLysSerAspLeuGlnGlnGlnPhe 171  
 Db 778 GGGGCTCCGGCGCCACGATCAACCTGATCCGCAAGAAACCTACCCATGAATTCAGAGGC 837  
 Qy 172 GlnThrArgIleGlyValHisGlySerLysLeuSerSerGlnGlyIleGlyTyrGlnVal 191  
 Db 838 CATGTCCGAACCTGGGC-----GGCGGAGTTGGAGCAATTACCGCTCC 879  
 Qy 192 GlyLysSerValAlaGly---ValSerGluAsnGlyAsnValIleuAlaArgLeuAspVal 210  
 Db 880 GAATCGACATCTACGCGGCCGACCTACGACCGATGCTCGGACACGTCGGCGCGCGGCTGACC 939  
 Qy 211 AspTyrArgThrThrGlyValAlaPheAspAlaAsnGlyLysArgIleAlaProGluPro 230  
 Db 940 GCCTACCAAGAGTAAAGCACTGTTCAATGAT----- 969  
 Qy 231 AlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsnThrAsnValAspTyrGln 250  
 Db 970 -----CACTACGAGCCCAAAACACAGCGTCTATTACGCGATC----- 1005  
 Qy 251 LeuAspAspLysGlnIleAsnIleAsnLeuAlaLeuThrHisTyrAsnAspLysGlnAspThr 270  
 Db 1006 CTCGAGTTCGATCGATGATCCGACACACCATGCTTACCTGCTCGCGCCGATATCAGACAAC 1065  
 Qy 271 AspTyrAlaProAspTyrGlyAsnArgLeuAlaValIlePheGlyGlyLysProSerLeu 290  
 Db 1066 GATCCCAAGGAGATCGGGCTGTGTGGGGCAGCTTCCGCTCTTGCAGACGCAAGGCAACGCC 1125  
 Qy 291 AsnAlaIle-----LysGlyLeuSerLeuSerGlnGlnProLysThrThr 305  
 Db 1126 AACGAGCTCCCGCTCTCTTCAACAACGAGCAAGTGAAGAGCTTGGAGACATGACAC 1185  
 Qy 306 LysSerThrPhe--AsnIleAsnTyrHisHisAspAspLeuTyrGlyAsnThrIle--- 323  
 Db 1186 CGCACGCTCTTGGCAACCTCGACCAATTTCCCAACGGCTGGGTGGCGCAAGGTGCA 1245  
 Qy 324 -----AsnThrAsnAlaTyrTyrArgArgGlyLysGlyArgPheTyrProPheVal 340  
 Db 1246 CTCGATCAACAAGATCAAGGCTTACACGGCGCCCTCGCGGCGATCATGGGCGACTGGCGG 1305  
 Qy 341 AlaProPheSerIleAlaLysValAlaLeuProIleLeuGlnSerMetAsnLeuProSerAla 360  
 Db 1306 GCACCGGACAAACAGCGCCAAAGATGCTTGGCGAGAAGTACACCGGCGAAACCAAGACAAAC 1365  
 Qy 361 ThrLeuAspAlaTyrThrLysAlaPro 369  
 Db 1366 TCGCTGGATATCTATCTCACCGGCGCC 1392

## RESULT 43

AAT95247  
 ID AAT95247 standard; DNA; 3438 BP.  
 XX  
 AC AAT95247;  
 XX  
 DT 14-APR-1998 (first entry)  
 XX  
 DE M. catarrhalis 4223 transferrin binding protein tbpA gene.  
 XX  
 KW Transferrin binding protein; tbpA; immunogen; vaccine; protection;  
 KW otitis media; antibody; diagnosis; therapy; carrier;  
 KW gene isolation; ss.  
 XX  
 OS Moraxella catarrhalis.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 214..3438  
 FT /\*tag= a  
 FT /product= transferrin\_binding\_protein  
 XX  
 PN WO9732980-A1.  
 XX  
 PD 12-SEP-1997.  
 XX  
 PF 07-MAR-1997; 97WO-CA00163.  
 XX  
 PR 03-JAN-1997; 97US-0778570.  
 PR 08-MAR-1996; 96US-0613009.  
 XX  
 PA (CONN-) CONNAUGHT LAB LTD.  
 XX  
 PI Du R, Harkness RE, Klein MH, Loosmore SM, Myers LE;  
 PI Schryvers AB, Yang Y;  
 XX  
 DR WPI; 1997-457533/42.  
 DR P-PSDB; NAW35312.  
 XX  
 PT DNA encoding transferrin receptor of a Moraxella strain - also  
 PT proteins, useful in vaccines, as diagnostic agents and in the  
 PT production of antibodies  
 XX  
 PS Claim 6; Fig 5; 162pp; English.  
 XX  
 CC The present sequence encodes the Moraxella catarrhalis 4223  
 CC transferrin binding protein tbpA, which can be used as an  
 CC immunogen, e.g. in vaccines to protect against diseases caused  
 CC by M. catarrhalis (specifically otitis media), or to raise  
 CC antibodies for diagnosis and therapy. It can also be used as a  
 CC carrier for other antigenic determinants, e.g. of bacteria  
 CC containing polysaccharide antigens or abnormal polysaccharides  
 CC present on tumour cells, particularly to make conjugate vaccines.  
 CC The tbpA DNA can be used to detect nucleic acid encoding  
 CC transferrin receptor protein, e.g. for diagnosis or gene  
 CC isolation, by usual hybridisation assays.  
 XX  
 SQ Sequence 3438 BP; 1138 A; 734 C; 726 G; 840 T; 0 other;  
 XX  
 Alignment Scores:  
 Pred. No.: 0.009 Length: 3438  
 Score: 134.00 Matches: 94  
 Percent Similarity: 35.32% Conservative: 48  
 Best Local Similarity: 23.38% Mismatches: 150  
 Query Match: 6.55% Indels: 110  
 Gaps: 21  
 DB:  
 US-09-889-746-2 (1-400) x AAT95247 (1-3438)  
 OY 10 LeuSerLeuProLeuSerValAla-----ValThrGlnGlnLeu 23  
 DB 277 TTGTCCTTTGGGTCTGCTTAACATCAGCGAGTGGCACTGGCAACACACGCCGATAAG 336

OY 24 TyrAlaGlnProAsnGluSerLeuProThrValGlnLeuGluProValValIleThrIle 43  
 DB 337 GCGGAGGCAACAGATAAGACAAACCTTGTGTTCTTGGATGAACCTGTTGTAAACGCG 396  
 OY 44 AspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrThrLysVal 63  
 DB 397 AAGAAAACGCCCGTAAA---GCCAACGAAGTTACAGGGCTTGGTAGGTGGTCAAACT 453  
 OY 64 IleTyrGluGluGlnIleGlnGluGlnAlaThrGlySerArgGlnLeuAlaAspValMet 83  
 DB 454 GCC---GAGACCATCAATAAAGAACAGTGTAAACATTCGAGACTTA-----498  
 OY 84 AlaGlnLeuIleProSerLeuGly---ValSerSerGlyThrThrSerAsnPheGlyGln 102  
 DB 499 ACAGCTATGACCTCGCATTTGCTGTGGTTCAGCAAGTCGTGGGCAAGCTCAGGCTAT 558  
 OY 103 ThrMetHisGly-----ArgGlnValGlnPheLeuAsnGlyVal-----116  
 DB 559 TCTATTCGTGTATGGATAAAATCGTGTGGCGGTATTGGTTGATGGCATCAATCAAGCC 618  
 OY 117 -----ProLeuThrGlySerArg---AspIleSerArgGlnLeu 128  
 DB 619 CAGCACTATGCCCTACAGGGCCCTGTGGCAGGCAAAATTTATCGCAGGTGGGCAATC 678  
 OY 129 AsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyAla---ThrSer 147  
 DB 679 AACGAATAGAAATACGAAATGTCCGCTCCGTTGAGATTAGTAAGGTGCAAAATTCAAAT 738  
 OY 148 IleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLysSer-----164  
 DB 739 GAATACGGCTCTGGGGCATTTATCTGGCTCTGTGCATTTGTATACCAAAACCCCGATGAC 798  
 OY 165 -----AspLeuGluGluGlnPheGluThrArg-----174  
 DB 799 ATCATCAAAAGATGTAAAGATTGGGGCGTGCAGACCAAAACCCGCTATGCCAGTAAAAAT 858  
 OY 175 -----IleGlyValHisGlySerLysLeuSerSerGlyGlyIle---187  
 DB 859 AACGCATGGTTAATTCTGTGGCAGCAGCAGGCAAGGCGGTCTCTTTAGCGGCTTATC 918  
 OY 188 -----GlyTyrGlnValGly 192  
 DB 919 ATCTACACCGCCCGCTGTCAGAAATACAGGCACATGATGCTATCAGGTAGC 978  
 OY 193 GlnSer-----ValAla 196  
 DB 979 CAAAGTTTGTAGAGCGTGGCAACCACTGACCCAAATAACCGAACATTTTAAATAGCA 1038  
 OY 197 GlyValSerGluAsnGlyAsnVal-----LeuAla 206  
 DB 1039 AATGAATGTGCCAATGGTAAATTATGAGCGGTGTGCTGGCGGTCAAAACCAACTTCAA 1098  
 OY 207 ArgLeuAspValAspTyrArgThrThrGlyGlyAlaPheAspAlaAsnGly---LysArg 225  
 DB 1099 GCCAAGCCCAACCAATGTCGTGATGAAGTCAATGTCAAGGATTATACAGGTCTTAACCGC 1158  
 OY 226 IleAlaProGluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsnThr 245  
 DB 1159 CTTATCCCCAAACCCCATC-----ACCCAAGACAGCAAAATCCTTACTGCTGCCCA 1209  
 OY 246 AsnValAspTrpGlnLeuAspAspLysGlnAsnIleAsn-----LeuAlaLeuThrHis 263  
 DB 1210 GGT-----TATCAGCTAAACGATAAGCACTATGTGCGTGGTGTGTATGAATACCAAA 1263  
 OY 264 TyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeu 283  
 DB 1264 CAAACTAGCCCATGCAAGATAAAACCGCTGCTGTAT-----CTGACGGTTTCAAT 1314  
 OY 284 PheGlyGluLysProSerLeuAsn-----AlaIleLysGlyLeu 296  
 DB 1315 GACATTGAAAAAATCAAGGCTCAGCAACCATGCCCAAGCAATGGCTATTATCAAGGCAAT 1374  
 OY 297 SerLeuSerGluGlnProLysThrThr-----LysSerThrPheAsnIleAsnTyr 313

```

Db 1375 AATCTTGGAAGCATGTCATGATACCATGGCCAGATTCAGATTATGGCATCACTAT 1434
Qy 314 H1SH1S 315
Db 1435 GCTCAT 1440

RESULT 44
AAS27475
ID AAS27475 strand: cDNA; 686 BP.
XX
AC AAS27475;
XX
DT 07-NOV-2001 (first entry)
XX
DE cDNA encoding novel signal transduction pathway protein, Seg ID 510.
XX
KW Neuroprotective; cytoskeletal; dermatological; immunosuppressive; tumour;
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KW immune system disorder; rheumatoid arthritis; inflammatory condition;
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW reproductive system; gastrointestinal; liver disorder; AIDS; ss;
KW acquired immune deficiency syndrome.
XX
OS Homo sapiens.
XX
XX WO200154733-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01312.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 26-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0217496.
PR 26-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 14-AUG-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 23-AUG-2000; 2000US-0227182.
PR 30-AUG-2000; 2000US-0227009.
PR 01-SEP-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.

```

```

PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0233397.
PR 14-SEP-2000; 2000US-0233398.
PR 14-SEP-2000; 2000US-0233399.
PR 14-SEP-2000; 2000US-0233400.
PR 14-SEP-2000; 2000US-0233401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.

```

```

Alignment Scores:
Pred. No.:          0.00113
Score:              135.50
Percent Similarity: 45.21%
Best Local Similarity: 25.00%
Query Match:        6.52%
DB:                  22
US-09-889-746-2 (1-400) x AAS27475 (1-686)
Length:             686
Matches:            47
Conservative:       38
Mismatch:           62
Indels:             41
Gaps:               9

```

PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226581.  
 PR 22-AUG-2000; 2000US-0226688.  
 PR 23-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231142.  
 PR 08-SEP-2000; 2000US-0231143.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0233423.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236602.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI; 2001-581633/65.  
 DR P-PSDB; AAB7504.  
 XX  
 PT New isolated nucleic acid encoding a protein for diagnosing,  
 PT preventing, treating or ameliorating medical conditions and used as  
 PT food additives or preservatives -  
 XX  
 PS Claim 1, SEQ ID No 424; 837bp; English.

The invention describes an isolated nucleic acid molecule (I) encoding a novel central nervous system protein. (I) and polypeptides (III) encoded by (I), are used to treat a medical condition and in diagnosis of a pathological condition. Disorders which are diagnosed or treated include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g. neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia, angioneurosis, nervous system disorders e.g. Alzheimer's disease and amyotrophic lateral sclerosis, infections caused by bacteria, viruses e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders e.g. corneal infection, gastrointestinal disorders e.g. dysphagia, adenocarcinomas and irritable bowel syndrome, reproductive system disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes and pituitary dwarfism, cancers and disorders at the cellular level e.g. leukaemia, disorders involving neovascularisation e.g. malignancies, respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g. acute kidney failure and blood related disorders e.g. myocardial infarction. The polypeptides can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues and in chemotaxis. The polypeptides can also be used as a food additive or preservative to increase or decrease storage capabilities, fat content, lipid, protein,

Alignment Scores:		
Pred. No.:	0.0013	686
Score:	135.50	47
Percent Similarity:	45.21%	Conservative: 38
Best Local Similarity:	25.00%	Mismatches: 62
Query Match:	6.52%	Indels: 41
DB:	23	Gaps: 9

US-09-889-746-2 (1-400) x ABK43834 (1-686)

Qy	14	LeuLeuSerValAlaValThrGlnGlnLeuTyAlaGlnProAsnGluSerLeuProThr	33
Db	109	TTGGTCAACTGGGGATTATGGGGTAGCCAGAGCCGACCATACTCTGTGT	168
Qy	34	ValGluLeuGluProValIleThrIleAspLysSerGlyMetAlaLeuAlaAsnArg	53
Db	169	TCACATGACGATACTATTGTCGTACC-----GCCGCCGAGCAG	207
Qy	54	IleThrGlnMetProHisThrThrLysValIleTyArgGluGlnIleGlnGluAla	73
Db	208	AACCTTAGAGCGCCTGGCGTTTCGACCATCAGCCGACATGAA---ATCGCAAAACCCG	264
Qy	74	ThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSer	93
Db	265	GTT---GCCCGCGATGTGTCGAGATCATC---CGTACCATGCCGCGGTAACTGTACC	318
Qy	94	SerGlyThrThrSerAsnPheGlyGlnThrMetHisGlyArgGlnValGln-----	110
Db	319	GGTAACTCCACCACT---GGTCAGCGTGGGAATAACCGACAGATTGATATTCGCGGT	372
Qy	111	-----PheLeuLeuAsnGlyValProLeuThr-----	119
Db	373	ATGGGTCGGGAAAACACGCTGATTTGATTGACGGCAAGCCGGTAAAGCAGCCGTAACTCG	432
Qy	120	-----GlySerArgAspIleSerArgGlnLeuAsnSerIleAsnPro	133
Db	433	GTGCGTCAGGCGCTGGCGTGGCGAGCGGATACCCGTGGTGATACTTCCTGGGTGCCACCT	492
Qy	134	AsnGlnValAlaArgIleGluValLeuSerGly---AlaThrSerIleTyArgLysGly	152
Db	493	GAATATGATTGAACGTATTGAAGTTCTCGGTGGTCCGCGACGTGCGGCTTATGCAACGCG	552
Qy	153	AlaThrGlyGlyLeuIleAsnIleValThrLysSerAspLeuGluGlnPheGlu	172
Db	553	GCGGGGCGCGGTGGTTAACTCATTTACCAAAAGAGCGAGCGCGAG-----	600
Qy	173	ThrArgIleGlyValHisGlySer	180



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 25, 2002, 20:36:34 ; Search time 52 Seconds

(without alignments)  
2359.052 Million cell updates/sec

Title: US-09-889-746-2

Perfect score: 2047

Sequence: 1 MRSHYFQMLSLPLSLAVT.....SKAEVLGRVPLNKKRALF 400

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DBV=x1h  
-O=/cg2\_1/USPTO SP001/US09889746/runat 20122002.143750.14338/app query.fasta\_1.583  
-DB=Issued\_Patents\_NA -QFMT=fastap -SUFFIX=eml -WINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=45  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09889746 @CGN 1.1 17 @runat 20122002.143750.14338 -ICPU=3  
-NO\_XIPXY -NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:\*  
1: /cg2\_6/prodata/1/ina/5A\_COMB.seq:\*  
2: /cg2\_6/prodata/1/ina/5B\_COMB.seq:\*  
3: /cg2\_6/prodata/1/ina/6A\_COMB.seq:\*  
4: /cg2\_6/prodata/1/ina/6B\_COMB.seq:\*  
5: /cg2\_6/prodata/1/ina/6CTUS\_COMB.seq:\*  
6: /cg2\_6/prodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	199.5	9.7	2495	4	US-09-668-113A-1 Sequence 1, App1
2	186	9.1	2182	4	US-09-221-017B-916 Sequence 916, App
3	178.5	8.7	4661	4	US-09-221-017B-970 Sequence 970, App
4	154.5	7.5	9057	4	US-09-453-702B-194 Sequence 194, App
5	149	7.3	37895	1	US-08-375-709-1 Sequence 1, App1
6	149	7.3	37895	1	US-08-375-709-1 Sequence 1, App1
7	149	7.3	37895	3	US-09-090-793-1 Sequence 1, App1
8	147.5	7.2	2910	1	US-08-375-709-6 Sequence 6, App1
9	147.5	7.2	2910	1	US-08-375-709-6 Sequence 6, App1
10	142	6.9	4534	4	US-09-221-017B-619 Sequence 619, App
11	138	6.7	8286	4	US-09-059-584-54 Sequence 54, App1
12	136	6.6	2091	1	US-08-765-081-4 Sequence 4, App1

13	136	6.6	2091	3	US-09-098-082-4	Sequence 4, App1
14	136	6.6	2091	5	PCT-US95-06394-4	Sequence 4, App1
15	136	6.6	8041	1	US-08-765-081-1	Sequence 1, App1
16	136	6.6	8041	1	US-09-098-082-1	Sequence 1, App1
17	136	6.6	8041	5	PCT-US95-06394-1	Sequence 1, App1
18	135	6.6	6737	4	US-09-453-702B-76	Sequence 76, App1
19	134	6.5	3222	3	US-08-613-009A-2	Sequence 2, App1
20	134	6.5	3222	4	US-08-778-570B-2	Sequence 2, App1
21	134	6.5	3222	4	US-09-059-584-2	Sequence 2, App1
22	134	6.5	3438	3	US-08-613-009A-1	Sequence 1, App1
23	134	6.5	3438	4	US-08-778-570B-1	Sequence 1, App1
24	134	6.5	3438	4	US-09-059-584-1	Sequence 1, App1
25	133	6.5	4651	3	US-08-425-843-6	Sequence 6, App1
26	130.5	6.4	7304	4	US-09-453-702B-174	Sequence 174, App
27	130	6.4	2169	5	PCT-US96-05320A-264	Sequence 264, App
28	130	6.4	3210	3	US-08-613-009A-6	Sequence 6, App1
29	130	6.4	3210	4	US-08-778-570B-6	Sequence 6, App1
30	130	6.4	3210	4	US-09-059-584-6	Sequence 6, App1
31	130	6.4	3660	3	US-08-613-009A-5	Sequence 5, App1
32	130	6.4	3660	4	US-08-778-570B-5	Sequence 5, App1
33	130	6.4	3660	4	US-09-059-584-5	Sequence 5, App1
34	130	6.4	87563	4	US-09-453-702B-57	Sequence 57, App1
35	128	6.3	1845	4	US-09-307-973A-1	Sequence 1, App1
36	128	6.3	2378	4	US-08-537-361E-7	Sequence 7, App1
37	128	6.3	2378	4	US-08-817-707-7	Sequence 7, App1
38	127	6.2	2760	4	US-09-221-017B-462	Sequence 462, App
39	123.5	6.0	4061	4	US-08-425-843-1	Sequence 1, App1
40	123	6.0	1844	4	US-09-307-973A-2	Sequence 2, App1
41	122	6.0	2376	4	US-08-537-361E-3	Sequence 3, App1
42	122	6.0	2376	4	US-08-817-707-3	Sequence 3, App1
43	121	5.9	2373	4	US-08-817-707-5	Sequence 5, App1
44	121	5.9	3537	2	US-08-363-124A-3	Sequence 3, App1
45	118	5.8	2800	3	US-08-448-194-3	Sequence 3, App1

#### ALIGNMENTS

RESULT 1  
US-09-668-113A-1  
Sequence 1, Application US/09668113A  
Patent No. 6410703  
GENERAL INFORMATION:  
APPLICANT: Russco, Thomas A.  
TITLE OF INVENTION: Identification of A Vaccine Candidate from an  
FILE REFERENCE: 11520.0214  
CURRENT APPLICATION NUMBER: US/09/668,113A  
CURRENT FILING DATE: 2000-09-22  
NUMBER OF SEQ ID NOS: 10  
SEQ ID NO 1  
LENGTH: 2495  
TYPE: DNA  
ORGANISM: Escherichia coli  
FEATURE:  
US-09-668-113A-1

Alignment Scores:  
Pred. No.: 5,12e-13  
Score: 199.50  
Percent Similarity: 41.94%  
Best Local Similarity: 24.63%  
Query Match: 9.75%  
DB: 4  
Gaps: 16

US-09-889-746-2 (1-400) x US-09-668-113A-1 (1-2495)

QY 9 TTPLeuSerLeuProLeuSerValAlaValThGInGLeu-----Ty 24  
DB 342 TGG---TTCGTAAGCTGTGCTCAAGTTGGGTTGAATAGCAGGTATCAGGCAATAC 398  
QY 25 AAGInProhngInuSerLeuProThrValGluLeuGluProValValIleThrIle 44  
DB 399 TCCGACGATGATATATGAC-----GAGACTCTGTGTGTGAA----- 434

```
QY 45 LysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrThrLysValIle 64
DB 435 -----GCCACCCCTGAGCAGGATTAAACACAGCAGCCGGCGTGTGCTATT 482
QY 65 TyrGluGluGlnIleGlnGluAlaThrGlySerArgGlnLeuAlaAspValMetAla 84
DB 483 ACCAGCGAGGATATT-----AAAAGACCCCTCCGTTAAACGACCTTTCAGATATTATTCGT 539
QY 85 GlnLeuIleProSerLeuGlyVal-----SerSerGlyThrThrSerAsnPhe 100
DB 540 AATATG---CCTGGTGTAACTTACCGCGCAATAGCGCTCGGCACACGCGGTAATAC 596
QY 101 GlyGlnThr-----MetHisGlyArgGlnValGlnPheLeuLeuAsnGlyVal 116
DB 597 CCGCAGATCGATATTCTGTGTGTATGCGGCCGGAACACACCTTAATTTTAAATGATGTGTA 656
QY 117 ProLeuThr-----GlySerArgAspIleSerArg 126
DB 657 CCGGTGACGTACGTAACCTCGTGTGCTTATAGTGGCGTGGGAGCGTATACCCCGGT 716
QY 127 GlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGly---Ala 145
DB 717 GACACCACTGGGTGCGCACCGGAACAGTTGAGCGTATTGAAGTATCGCGCGCCCTGCG 776
QY 146 ThrSerIleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLysSerAsp 165
DB 777 GCGGCGCGTACGTTTCCGGGGCGCGCGGGGGTGTGAACATCATTAACCAACGTCC 836
QY 166 LeuGluGluGlnPheGluThrArgIleGlyValHisGlySerIleLysLeuSerGlu 185
DB 837 ACCAAGACTGGGACGGTTCGCTGCTGTTATACCAACACGCGGGAAGTAGCGAAG 896
QY 186 GlyIleGlyTyrGlnValGlyGlnSerValAlaGlyValSerGluAsnGlyAsnValLeu 205
DB 897 GCGGTACGCGTCCGCCAATTTACGCTTAGTGGCCTCTGCTGGTGTATCTCTTACC 956
QY 206 AlaArgLeuAspValAspTyr---ArgThrThrGlyGlyAlaPheAspAlaAsnGlyLys 224
DB 957 ACGCGTTTGTATGTTAACTGTAATTAACCGGATGCTGACAGTTGGGATTAAT----- 1010
QY 225 ArgIleAlaProGluProAlaGlnThrAspLys-----GlnAspSerLysSer 240
DB 1011 -----TCTCCGTGCGTACGAAACGCGCGCGCATGAGGGGTACGTATAC 1058
QY 241 LeuSerValAsnThrAsnValAspTyrGlnLeuAspAspLysGlnAsnIleAsnLeuAla 260
DB 1059 AAGATATTACCGCGTGTCTCGTGGAAATTAATCCGACGAGATTCTCGATTTCGAA 1118
QY 261 LeuThrHisTyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeu 280
DB 1119 GTCCGATAT-----AGCGCGCAGGGAATATCTATGCGGGCGATACGACGACAGTTCT 1172
QY 281 AlaValLeuPheGlyGluLysProSerLeuAsnAlaIleLysGlyLeuSerLeuSerGlu 300
DB 1173 TCCAGTGCAGTTACCGAA-----AGCCTGGCAAAA 1202
QY 301 GlnProLysThrThrLysSerThrPheAsnIleAsnTyr-----HisHisAspAspLeu 318
DB 1203 TCCGCAAGACGACGACCCCTGTACCGACAGAAATATGGCATTACGCATTAATGGTATC 1262
QY 319 Trp 319
DB 1263 TGG 1265
RESULT 2
US-09-221-017B-916/c
; Sequence 916, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221.017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P11182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: P2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Morroy, Gladys H.
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 916:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2182 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORGANISM: PORPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1...2182
; US-09-221-017B-916
Alignment Scores:
Pred. No.: 1,53e-11 Length: 2182
Score: 186.00 Matches: 78
Percent Similarity: 39.76% Conservative: 52
Best Local Similarity: 23.85% Mismatches: 105
Query Match: 9.09% Indels: 92
DB: 4 Gaps: 15
US-09-889-746-2 (1-400) x US-09-221-017B-916 (1-2182)
QY 25 AlaGlnProAsnGluSerLeuProThr-----ValGluLeuGluProValIleThr 42
DB 1951 GCCCAACCTACCGACACCATCGTATCCGCAATATCGCACTTGAGGATATAGTGTGACC 1892
QY 43 IleAspLysSerGly-----MetAlaLeuAlaAsnArgIleThrGlnMetPro 58
DB 1891 GGTAGCGTACAGCCCGCTGCTTAAAGATGATACCTGCCCAAAAGGTGTTCAAGGCCA 1832
QY 59 HisThrThrLysValIleTyrGluGluGlnIleGlnGluAlaThrGlySerArgGln 78
DB 1831 AAGGATATCAAGCTATA-----GCCCATCTTCT 1802
```

```

Qy 79 LeuAlaSerValMetAlaGlnLeuIleProSerLeuGlyValSerSerClyThrThrSer 98
Db 1801 TTCAATTGACGTAAGTACGAGTATATTCTTCCGGGATCGAGTT----- 1760
Qy 99 AasnheGlyGlnThrMetHisGlyArgGlnValGln----- 110
Db 1759 -----ACCAAGCATGGTCCAGAGATCAAGTCAATGCTCAAGGATTTGACGAA 1712
Qy 111 -----PheLeuLeuAasnGlyValProLeuThrClySerArgAspIleSerArg 126
Db 1711 AGTTCTATTCTCTTCCTGCGTCGATGGGAATTGATTTCACAGCGGATTCACAGTGAATA 1652
Qy 127 GlnLeuAasnSerIleAasnProAasnGlnValAlaArgIleGluValLeuSerGlyValAlaThr 146
Db 1651 GACTTCGACAGCAATCAATCCGATGATGATCATCGACGCAATCGAACTGCTGTCGACCTTCC 1592
Qy 147 Ser--IleTyrGlySerGlyAlaThrGlyGlyLeuIleAasnIleValAlaThrLysSerAsp 165
Db 1591 TCTCTTTGTATCGGATCTAATGATCCATCGAGGTGTTATCAATATCATCACCCGTACA--- 1535
Qy 166 LeuGlnGlnGlnGlnPheGluThrArgIleGlyValHisGlySerLysLeuSerSerGlu 185
Db 1534 ---GCCAAGGATCTTTT-----CGGTAATCTGCTTCGCTCGATACGATAGTGGCAGC 1484
Qy 186 GlyIleGlyTyrGlnValGlyGlnSerValAlaGlyValSerGlyAasnGlyAasnValLeu 205
Db 1483 GAGCAGAAATACGATCTG-----GCAGCAGAGAGT---AAACGTGGATCTTCAAC 1436
Qy 206 AlaArgLeuAepValAaspTyrArgThrThrGlyGlyAlaPheAspAlaAasnGlyLysArg 225
Db 1435 AGTCAGTCCGGCGTCGACAAATATAGACAGACAAAGAATTATATTTGGCCGAT----- 1385
Qy 226 IleAlaProGluProAlaGlnThrAepLysGlnAAspSerLysSerLeuSerValAasnThr 245
Db 1384 -----CAGTTTCAGACAGAACTGAATTTTCAGGC 1355
Qy 246 AasnValAepTyrGlnLeuAepAspLys-----GlnAasnIleAasnLeu 259
Db 1354 AATTCAGCATGGAATTAATATCAGAGATTACAATCTCCACGSAANAATTATCTTC 1295
Qy 260 AlaLeuThr-----HisTyrAasnAepLysGlnAAspThrAsp 271
Db 1294 AACCTTACCGGACGTGCAATCTGCGTAAACGACACCTGACGGATTAAGTTCATTTCTG 1239
Qy 272 TyrAlaProAepTyrGlyAasnArgLeuAlaValLeuPheGlyGlyLysProSerLeuAasn 291
Db 1234 TAC-----AACTCTTACGATGTCAA 1214
Qy 292 AlaIleLysGlyLeuSerLeuSerGlnGlnProLysThrThrLysSerThrPheAasnIle 311
Db 1213 GCCGGTGCACACTGGCGATACAGGAG-----ACTTCGATTCGACAGCTC 1165
Qy 312 AsnTyrHisIleAepAaspLeu 318
Db 1168 TCTTATCATTTACGACAAATATA 1148

RESULT 3
US-09-221-017B-970
; Sequence 970, Application US/09221017B
; Patent No. 644479
; GENERAL INFORMATION:
; APPLICANT: Rose, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USRS THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:

```

```

: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: Windows
: SOFTWARE: FASTSEQ for Windows Version 2.0.b
: CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/09/221.017B
:   FILING DATE: 23-DEC-1998
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: PP1182
:   FILING DATE: 31-DEC-1997
: PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: PP1546
:   FILING DATE: 30-JAN-1998
: PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: PP2911
:   FILING DATE: 09-APR-1998
: PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: PCT/AU98/01023
:   FILING DATE: 10-DEC-1998
: ATTORNEY/AGENT INFORMATION:
: NAME: Montoy, Gladys H
: REGISTRATION NUMBER: 32,430
: REFERENCE/DOCKET NUMBER: 27340-20021.00
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 650-813-5600
: TELEFAX: 650-494-0792
: TELEX: 706141
: INFORMATION FOR SEQ ID NO: 970:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4661 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: circular
: MOLECULAR TYPE: DNA (genomic)
: HYPOTHEetical: NO
: ANTI-SENSE: UNKNOWN
: ORIGINAL SOURCE:
: ORGANISM: PORPHYROMONAS GINGIVALIS
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: 1...4661
US-09-221-017B-970

Alignment Scores:
Pred. No.:          3,94e-10      Length:         4661
Percent Similarity: 178.50        Matches:           83
Best Local Similarity: 25.08%     Mismatches:       56
Query Match:        8.72%         Indels:            45
DB:                  4             Gaps:              16

US-09-889-746-2 (1-400) x US-09-221-017B-970 (1-4661)
QY      31 LeudProthrValGluLeugluproValIleThrlAsplySerglyMetAlaleu 50
      ||| |||| |
Db    2267 CTGGCAAGCAAACACTTGGAAGTCGTGCATTACC-----GTTACCGGTACA 2314
      ||| |||| |
QY      51 AlaanaarglierghinmetProhistrThrlyValIletyrgluglnlglin 70
      ||| |||| |
Db    2315 CGTAAcgcgctggcgatgcgtcccttgaggcaacgaagtccttaaccgttaaagacaATA-- 2371
      ||| |||| |
QY      71 GluGlinaIathrglySerArgInleuaLaaspValmetAlagin---Leuilleproser 89
      ||| |||| |
Db    2372 -----GCCCTTCTTCGGCCTCCACTCTCGAAGCCTTAATGCAGAAGGCGTAGTCGCT 2425
      ||| |||| |
QY      90 LeuglyvalberSerglyThrSerasnPheglyInThr----MethisglYArg 107
      ||| |||| |
Db    2426 TTtGACTTCGCCCCCATCTGAATGGCGCTCTTCAATGCAAGTGAACGGCCTTAGCAGTAAG 2485
      ||| |||| |
QY      108 GluValGINpheLeuleuauenglyValproLeuthrglySerArgAspliserargin 127
      ||| |||| |
Db    2486 TATATCTCATCTTATATCGAATGTAAGCGGTGTACGSGGAGATGTAAGCCGTCAGCCGAT 2545
```





GENERAL INFORMATION:  
APPLICANT: YAZAWA, Kazunaga  
APPLICANT: YAMADA, Akiko  
APPLICANT: KATO, Seishi  
APPLICANT: KONDO, Kiyosi  
TITLE OF INVENTION: Gene Coding for Eicosapentaenoic Acid Synthesizing  
TITLE OF INVENTION: Enzymes and Process for Production of Eicosapentaenoic  
TITLE OF INVENTION: Acid  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/752,929  
FILING DATE: 20-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/375,709  
FILING DATE: 20-JAN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/178,251  
FILING DATE: 14-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 4-147945  
FILING DATE: 15-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: WEGNER, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 53466/150/AAOK  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 37895 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: Shewanella putrefaciens SCRC-2874 (FERM  
ORGANISM: BP-1625)  
US-08-752-929-1  
Alignment Scores:  
Pred. No.: 3.28e-05 Length: 37895  
Score: 149.00 Matches: 64  
Percent Similarity: 46.03% Conservative: 52  
Best Local Similarity: 25.40% Mismatches: 83  
Query Match: 7.28% Indels: 54  
DB: 12 Gaps: 1  
US-09-889-746-2 (1-400) x US-08-752-929-1 (1-37895)  
Qy 3 HisSerHisThrPheGlnThrLeuSerLeuProLeu----- 14  
Db 9660 TATAACAAATATATTAGGAATGAGTATGTTTAAATCAAACTTCCGGCTCAGTC 9719  
Qy 15 -----LeuSerValAlaValThrGlnGlnLeuTyAlaGlnPro-----Aan 28  
Db 9720 AAACCTTGCCATATCCGACGCTTAACAGCTCGCTA--GCTATGCTGTTTTTGCAGAA 9776  
Qy 29 GluSerLeuProThrValGluLeuGluProValValleThrIleAspLysSerGlyMet 48

Db 9777 GAAACTGCTGCTGAAGAACAAATAGAAAGAGTCGACGATGACC-----GGATCG 9824  
Qy 49 AlaLeuAlaAsnArgIleThrGlnMetProHisThrThrIleValIleTyGluGluGln 68  
Db 9825 CGAATCGTTAAAGCAGAGCTAACTCAACAGCTCCAGTCGTCAGCCCTTTCAGCCGAGAA 9884  
Qy 69 IleGlnGluGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIlePro 88  
Db 9885 CTGACAAA-----TTTGGTAATCAAGATTAGGTAGCGTACTAGCAGAAATTA---CCT 9935  
Qy 89 SerLeuGlyVal-----SerSerGlyThr 96  
Db 9936 GCTATTGGTGCACCAACACTATTATTGGTAAATAACATAGCAACTCAAGCGCAGGTGTT 9995  
Qy 97 ThrSerAsnPheGlyGlnThrMetHisGlyArgGlnValGlnPheLeuLeuAsnGly--- 115  
Db 9996 AGCTCAGCAGACTTTCGCTCGTCTAGGTGCTAACAGAACCTTAGTATTAGTCAACGGTAAG 10055  
Qy 116 -----ValProLeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAen 132  
Db 10056 CGCTACGTTGCCGCGCAACCGGCTCAGCTGAGTA-----GATTGTCAACTATACCA 10109  
Qy 133 ProAsnGlnValAlaArgIleGluValLeuSer---GlyAlaThrSerIleTyGlySer 151  
Db 10110 ACTAGCATGATCTCGCGAGTTGAGATTGTAACCGCGGTGCTTCAGCAATTTATGGTTTCG 10169  
Qy 152 GlyAlaThrGlyLeuIleAsnIleValThrLysSerAspLeuGluGluGlnPhe 171  
Db 10170 GACGCTGTATCAGGTGTTATCAAGTTATCTTAAAGAGACTTTGAAGGCTTTGAGTTT 10229  
Qy 172 GluThrArgIleGlyValHisGlySerLysLeuSerSerGluGlyIleGlyTyGlnVal 191  
Db 10230 AACGCACGT-----ACTAGCGGTCTCTACTAAAGTGTAGGCACCTCAAGAG 10274  
Qy 192 -----GlyGlnSerValAlaGlyValSerGluAsnGlyAsnVal 204  
Db 10275 CACTCTTTTGGACATTTGGGTGGTGCACAACTTGCAGATGACGCTGTAATGT-AACCTT 10333  
Qy 205 LeuAlaArgLeuAspValAspTyArgThrThrGly 216  
Db 10334 CTACGACGTTATGACGTACAAAGAAAGTCAATGCC 10369  
RESULT 7  
US-09-090-793-1  
Sequence 1, Application US/09090793  
Patent No. 6140486  
GENERAL INFORMATION:  
APPLICANT: Calgene, LLC  
TITLE OF INVENTION: Production of polyunsaturated fatty acids by expression  
FILE REFERENCE: CGNE.131.01US  
CURRENT APPLICATION NUMBER: US/09/090,793  
EARLIER FILING DATE: 1998-06-04  
EARLIER APPLICATION NUMBER: 60/048,650  
EARLIER FILING DATE: 1997-06-04  
NUMBER OF SEQ ID NOS: 66  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 1  
LENGTH: 37895  
TYPE: DNA  
ORGANISM: Shewanella putrefaciens  
US-09-090-793-1  
Alignment Scores:  
Pred. No.: 3.28e-05 Length: 37895  
Score: 149.00 Matches: 64  
Percent Similarity: 46.03% Conservative: 52  
Best Local Similarity: 25.40% Mismatches: 83  
Query Match: 7.28% Indels: 54  
DB: 12 Gaps: 1  
US-09-889-746-2 (1-400) x US-09-090-793-1 (1-37895)

```

Qy 3 HisSerHisTyrPheGlnIleProLeuSerLeuProLeu----- 14
Db 9660 TATACAAATTATATTAAGGAATAGATATGTTTAAATCAAAACTTTCGGCTCAGTC 9719
Qy 15 -----LeuSerValAlaValThrGlnGlnLeuTyrAlaGlnPro-----Aan 28
Db 9720 AAACCTGCATATCCCGACAGCTTACAGCCCTGCTA---GCTATAGCTCTTTTTCAGAA 9776
Qy 29 GluSerLeuProThrValGlnLeuGluProValIleThrIleAspLysSerGlyMet 48
Db 9777 GAAATCGCTGCTGAAGAACAAATAGAAAGTCCAGTGACC-----GATTCG 9824
Qy 49 AlaLeuAlaAsnArgIleThrGlnMetProHisThrTyrValIleTyrGlnGln 68
Db 9825 CGAATCGCTAAAGACAGCTAAGTCAACAGCTCCAGTGTGACCTTTCAGCCGAAAGAA 9884
Qy 69 IleGlnGlnGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIlePro 88
Db 9885 CTGACACAAA-----TTTGTATATCAAGATTAGTACGCTACTACAGAAATTA---CCT 9935
Qy 89 SerLeuGlyVal-----SerSerGlyThr 96
Db 9936 GCTATTGGTGCAACCAACTATTATTGTTAATACAAATAGCAACTCAAGCGAGGTGT 9995
Qy 97 ThrSerAsnPheGlyGlnThrMetHisGlyArgGlnValGlnPheLeuAsnGly--- 115
Db 9996 AGCTCAGCAGACTGTGCTGTCTAGGTGCTACAGAACTTATGATTAAGCAAGCTAAG 10055
Qy 116 -----ValProLeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAsn 132
Db 10056 CGCTACGTTGCCGCGCAACCGGCTCAGCTCAGGTGTA-----GATTGTCAACTATACCA 10109
Qy 133 ProAsnGlnValAlaArgIleGlyValLeuSer---GlyAlaThrSerIleTyrGlySer 151
Db 10110 ACTAGCATATCTCGCGAGTGAATTGTAACCGCGGCTTCACAGCAATTATGATGCTG 10169
Qy 152 GlyAlaThrGlyGlyLeuIleAsnIleValThrLysSerAspLeuGlnGlnGlnPhe 171
Db 10170 GACGCTGATCAGCGTGTATCAACGTTATCCTTAAAGAAAGCTTGAAGCTTGAAGTT 10229
Qy 172 GluThrArgIleGlyValHisGlySerLysLeuSerSerGlyIleGlyTyrGlnVal 191
Db 10230 AACGCAAGT-----ACTAGCGGCTTACTGAAAGTGTAGGCACTCAAGAG 10274
Qy 192 -----GlyGlnSerValAlaGlyValSerGluAsnGlyAsnVal 204
Db 10275 CACTCTTTTGACATTTTGGGTGGTGCAAAAGCTTGACAGATGAGCTGTATATGT-AACTT 10333
Qy 205 LeuAlaArgLeuAspValAspTyrArgThrThrGly 216
Db 10334 CTACGCAAGTTATGAACGTAACAAAGAGTCATGGC 10369

RESULT 8
; Sequence 6, Application US/08375709
; Patent No. 5683898
; GENERAL INFORMATION:
; APPLICANT: YAZAWA, Kazunaga
; APPLICANT: YAMADA, Akiko
; APPLICANT: KATO, Seishei
; APPLICANT: KONDO, Kiyoshi
; TITLE OF INVENTION: Gene Coding For Bicosapentaenoic Acid
; TITLE OF INVENTION: Synthesizing Enzymes and Process for Production of
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/375,709
; FILING DATE: 20-JAN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/178,251
; FILING DATE: 14-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 4-147945
; FILING DATE: 15-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: WEGNER, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 53466/150/AAOK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2910 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Shewanella putrefaciens SCRC-2874 (FERM
; ORGANISM: BP-1625)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2910
; NAME/KEY: mat peptide
; LOCATION: 1..2910
; US-08-375-709-6

Alignment Scores:
Pred. No.: 7,31e-07 Length: 2910
Score: 147.50 Matches: 58
Percent Similarity: 47.79% Conservative: 50
Best Local Similarity: 25.66% Mismatches: 78
Query Match: 7.21% Indels: 41
DB: 1 Gaps: 10

US-09-889-746-2 (1-400) x US-08-375-709-6 (1-2910)
Qy 15 LeuSerValAlaValThrGlnGlnLeuTyrAlaGlnProAsnGluSerLeuProThrVal 34
Db 61 TTACAGCCTCGCTGCTAGTGCCTGTTTTCAGAA-----GAAACTGCTGCTGAAGAA 114
Qy 35 GluLeuGluProValIleThrIleAspLysSerGlyMetAlaLeuAlaAsnArgIle 54
Db 115 CAATAGAAAGAGTGCAGTAC-----GATCGCGAATGCTTAAAGCAGAG 162
Qy 55 ThrGlnMetProHisThrThrLysValIleTyrGlnGlnGlnIleGlnGlnAlaThr 74
Db 163 CTAACTCAACCAAGCTCCAGTGTGACGCTTTCAGCGGAAAGAACTGACAAA-----TTT 216
Qy 75 GlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyVal----- 92
Db 217 GGTAATCAAGATTAGTACGCTACTACAGCAATTA---CTGCTATTGTCGAACCAAC 273
Qy 93 -----SerSerGlyThrThrSerAsnPheGlyGln 102
Db 274 ACTATTATTGTTAATACAAATAGCAACTCAAGCGAGGTGTTAGCTCAGCAAGCTTGGCT 333
Qy 103 ThrMetHisGlyArgGlnValGlnPheLeuLeuAsnGly-----ValProLeu 118

```

```

Db 334 CGTCTAGGTGTAACAGAACCTTAGTATTAGTCAACGGTAAGCGCTACGTTCCGGCCAA 393
Qy 119 ThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArg 138
Db 394 CCGGGCTCAGCTGAGTA-----GATTTGTCAACTATACCAACTAGCATGATCTCGCGA 447
Qy 139 IleGluValLeuSer---GlyAlaThrSerIleTyGlySerGlyAlaThrGlyGlyLeu 157
Db 448 GTTGAGATTGTAACCGCGGCTCTTACGCAATTTATGTTGCGACGCTGTATCAGGTGTT 507
Qy 158 IleAsnIleValThrLysSerAspLeuGluGlnPheGluThrArgIleGlyVal 177
Db 508 ATCAACGTTATCTTAAAGAACCTTTGAGCGCTTGGAGTTTAAAGCGAGT----- 558
Qy 178 HisGlySerLysLeuSerSerGluGlyIleGlyTyGlnVal----- 191
Db 559 -----ACTAGCGGTTCTACTGAAAGTGTAGGCACTCAAGAGCACTCTTTTGACATTG 612
Qy 192 ---GlyGlnSerValAlaGlyValSerGluAsnGlyAsnValLeuAlaArgLeuAspVal 210
Db 613 GGTGGTGCAACGTTGCAGATCGACGTGGTAATGT-AACCTTCTACGCGAGGTTATGAACG 671
Qy 211 AspTyrArgThrGly 216
Db 672 TACAAAGAGTCATGCG 689

RESULT 9
US-08-752-929-6
; Sequence 6, Application US/08752929
; Patent No. 5798259
; GENERAL INFORMATION:
; APPLICANT: YAZAWA, Kazunaga
; APPLICANT: YAMADA, Akiko
; APPLICANT: KATO, Seishi
; APPLICANT: KONDO, Kiyosi
; TITLE OF INVENTION: Gene Coding for Eicosapentaenoic Acid Synthesizing
; TITLE OF INVENTION: Enzymes and Process for Production of Eicosapentaenoic
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESS: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/752,929
; FILING DATE: 20-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/375,709
; FILING DATE: 20-JAN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/178,251
; FILING DATE: 14-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 4-147945
; FILING DATE: 15-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: WEGNER, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 53466/150/AAOK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 6:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 2910 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Shewanella putrefaciens SCRC-2874 (FERM
; ORGANISM: BP-1625)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2910
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1..2910
US-08-752-929-6

Alignment Scores:
Pred. No.: 7,31e-07 Length: 2910
Score: 147.50 Matches: 58
Percent Similarity: 47.79% Conservative: 50
Best Local Similarity: 25.66% Mismatches: 78
Query Match: 7.21% Indels: 41
DB: 1 Gaps: 10

US-09-889-746-2 (1-400) x US-08-752-929-6 (1-2910)
Qy 15 LeuSerValAlaValThrGlnGlnLeuTyrAlaGlnProAsnGluSerLeuProThrVal 34
Db 61 TTAACAGCCTCGCTAGCTATGCTGCTTTTTCAGAA-----GAACTGCTGCTGAAGAA 114
Qy 35 GluLeuGluProValIleThrIleAspLysSerGlyMetAlaLeuAlaAsnArgIle 54
Db 115 CAAATAGAAAGAGTGCAGTGAC-----GGATCGGAATCGCTAAAGCAGAG 162
Qy 55 ThrGlnMetProHisThrThrLysValIleTyrGluGlnGlnIleGlnGluAlaThr 74
Db 163 CTAACCAACAGCTCCAGTCGTGAGCTTTCAGCCGGAAGAACTGACAA--TTT 216
Qy 75 GlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyVal----- 92
Db 217 GGTATCAAGATTAGGTAGCTACTAGCAGAAATTA---CCTGCTATTGGTGAACCAAC 273
Qy 93 -----SerSerGlyThrThrSerAsnPheGlyGln 102
Db 274 ACTATTATTGGTAATAACAATAGCAACTCAAGCGCAGGTGTAGCTCAGCAGACTTGGT 333
Qy 103 ThrMetHisGlyArgGlnValGlnPheLeuLeuAsnGly-----ValProLeu 118
Db 334 CGTCTAGGTGCTAACAGAACCTTAGTATTAGTCAACGGTAAGCGCTAGCTTCCGGCCAA 393
Qy 119 ThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArg 138
Db 394 CCGGCTCAGCTGAGTA-----GATTTGTCAACTATACCAACTAGCATGATCTCGCGA 447
Qy 139 IleGluValLeuSer---GlyAlaThrSerIleTyGlySerGlyAlaThrGlyGlyLeu 157
Db 448 GTTGAGATTGTAACCGCGGCTCTTACGCAATTTATGTTGCGACGCTGTATCAGGTGTT 507
Qy 158 IleAsnIleValThrLysSerAspLeuGluGlnPheGluThrArgIleGlyVal 177
Db 508 ATCAACGTTATCTTAAAGAACACTTTGAGCGCTTTGAGCGCTTTTAAAGCGAGT----- 558
Qy 178 HisGlySerLysLeuSerSerGluGlyIleGlyTyGlnVal----- 191
Db 559 -----ACTAGCGGTTCTACTGAAAGTGTAGGCACTCAAGAGCACTCTTTTGACATTG 612
Qy 192 ---GlyGlnSerValAlaGlyValSerGluAsnGlyAsnValLeuAlaArgLeuAspVal 210
Db 613 GGTGGTGCAACGTTGCAGATCGACGTGGTAATGT-AACCTTCTACGCGAGGTTATGAACG 671
Qy 211 AspTyrArgThrGly 216

```



Db 672 TACAAAAGATCATGCC 689

RESULT 10

US-09-221-017B-619

Sequence 619, Application US/09221017B

Patent No. 6444799

GENERAL INFORMATION:

APPLICANT: Ross, Bruce C.

TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF

NUMBER OF SEQUENCES: 1120

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FORSTER

STREET: 755 PAGE MILL ROAD

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows

SOFTWARE: FASTSEQ for Windows Version 2.0b

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/221,017B

FILING DATE: 23-DEC-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1182

FILING DATE: 31-DEC-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1546

FILING DATE: 30-JAN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP2911

FILING DATE: 09-APR-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU98/01023

FILING DATE: 10-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Montoy, Gladys H.

REGISTRATION NUMBER: 32,430

REFERENCE/DOCKET NUMBER: 27340-20021.00

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-813-5600

TELEFAX: 650-494-0792

TELEX: 706141

INFORMATION FOR SEQ ID NO: 619:

SEQUENCE CHARACTERISTICS:

LENGTH: 4534 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: UNKNOWN

ORIGINAL SOURCE:

ORGANISM: PORYPHYROMONAS GINGIVALIS

FEATURE:

NAME/KEY: misc\_feature

LOCATION: 1...4534

US-09-221-017B-619

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best Local Similarity:

Query Match:

DB:

US-09-889-746-2 (1-400) x US-09-221-017B-619 (1-4534)

QY 14 LeuleuserValaValThrGInGInLeuTyAlaGIn-----ProAsnGlu 29

Db 4094 CTACTATGATGTTGATTTACAGCGAGCTTTTGCACAGACGACATCCGACAGAT 4153

QY 30 SerLeuPProThValGInLeuGInProValValIleThrIleAspLysSerGlyMetAla 49

Db 4154 TCGCTACGATACACAACTTTACAGCCGTACAGGCTTATTACACGACGCGCTACT 4213

QY 50 LeuAlaAsnArgIleThrGInMetProHsrThrTrpValIleTyGInGInIle 69

Db 4214 CTG-----AAAAATACCGGACAGATGAGAACTCATCTCATCGGCAACATC 4261

QY 70 GInGInGInAlaThrGlySerArgGInLeuAlaAspValMetAlaGInLeuIleProSer 89

Db 4262 AAGCAG-----TCCGGCTTTAAACAACATGACCGCATCTCC---AAGCGCAAGTTG 4312

QY 90 LeuGlyValSerSer-----GlyThrTrpSerAsnPhe-----GlyGInThrMet 104

Db 4313 CTGATGATCATCAATACCGGCGCTTTAGTTGAAACATGCTATCCGCGTTCAAGCCC 4372

QY 105 HisGlyArgGInValGInPheLeuLeuAsnGlyValProLeuThrGlySerArgAspIle 124

Db 4373 TCCGGCAAGTATGTAACGCTATTGGTAAACGCGATCCCT---GCGGAAACGACATATC 4429

QY 125 SerArgGInLeuAsnSerIleAsnProAsnGInValAlaArgIleGlyValLeuSerGly 144

Db 4430 TCT-----ACGCTCAACACGACGCAACATCGAACAATCGAGATCTCTCAAGGC 4477

QY 145 Ala-----ThrSerIleTyGInSerGlyAlaThrGlyValLeuIleAsnIle 160

Db 4478 CCGTTCTTCTCATCTACCGCACCAATGCCATGCGGGGTGTGTGAACAT 4528

RESULT 11

US-09-059-584-54

Sequence 54, Application US/09059584

Patent No. 6440701

GENERAL INFORMATION:

APPLICANT: Myer, Lisa E

APPLICANT: Schryvers, Anthony B

APPLICANT: Hartness, Robin E

APPLICANT: Loosmore, Sheena M.

APPLICANT: Du, Run-Pan

APPLICANT: Yang, Yan-Ping

APPLICANT: Klein, Michael H

TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella

NUMBER OF SEQUENCES: 60

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sim & McBurney

STREET: 6th Floor, 330 University Avenue

CITY: Toronto

STATE: Ontario

COUNTRY: Canada

ZIP: M5G 1R7

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/059,584

FILING DATE: 14-APR-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/778,570

FILING DATE: 03-JAN-1997

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I

REGISTRATION NUMBER: 24973

REFERENCE/DOCKET NUMBER: 1038-794

TELECOMMUNICATION INFORMATION:

TELEPHONE: (416) 595-1155

TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 54:

SEQUENCE CHARACTERISTICS:  
LENGTH: 8266 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-059-584-54

Alignment Scores:  
Pred. No.: 5,148-05 Length: 8266  
Score: 138.00 Matches: 95  
Percent Similarity: 35.57% Conservative: 48  
Best Local Similarity: 23.63% Mismatches: 149  
Query Match: 6.74% Indels: 110  
Gaps: 21

US-09-889-746-2 (1-400) x US-09-059-584-54 (1-8266)

```
QY 10 LeuSerLeuProLeuSerValAla-----ValThrGlnGlnLeu 23
Db 133 TTGTCCTTTGGGCTGCTTAACATCATCGGAGGTGGCACTGGCAACACACGCGCGATAG 192
QY 24 TyrAlaGlnProAsnGlnSerLeuProThrValGluLeuGluProValValIleThrIle 43
Db 193 GCGGAGCAACAGATAGACAAACCTTGTGTCTTGGATGAAACTGTTGTAACAGCG 252
QY 44 AspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProIleThrThrLysVal 63
Db 253 AAGAAACCGCCGCTAAA---GCCAACGAAGTTACACGGGCTTGGTAAGGTGTCAAAAC 309
QY 64 IleTyrGluGluGlnIleGlnGluGlnAlaThrGlySerArgGlnLeuAlaAspValMet 83
Db 310 GCC---GAGACCATCAATAAGAACCAAGTGTAAACATTCGAGACTTA----- 354
QY 84 AlaGlnLeuIleProSerLeuGly---ValSerSerGlyThrThrSerAsnPheGlyGln 102
Db 355 ACACGCTATGACCTGGCATGTGCTGGTGTAGCAAGGTGCTGGGCAAGCTCAGGCTAT 414
QY 103 ThrMetHisGly-----ArgGlnValGlnPheLeuLeuAsnGlyVal----- 116
Db 415 TCTATTCTGGTGTGATATAAAATCGTGTGGCGGTATGTTGATGGCATCAATCAAGCC 474
QY 117 -----ProLeuThrGlySerArg---AspIleSerArgGlnLeu 128
Db 475 CAGCACTATGCCCTACAGCCCTGTGGCAGGCAAAAATTTATGCCCGAGTGGGCAATC 534
QY 129 AsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyAla---ThrSer 147
Db 535 AACGAAATAGATAACGAAATGTCCGCTCGTTGAGATTAGTAAAGGTGCAAAATTCAGCT 594
QY 148 IleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLysSer----- 164
Db 595 GAATACGGCTCTGGGCGATTATCTGGCTCTGGCATTTTGTACCAAAACCGCGATGAC 654
QY 165 -----AspLeuGluGluGlnPheGluThrArg----- 174
Db 655 ATCATCAAGATGGTAAAGATTGGGCGGTGGCAGCAACCAACCGCTATGCCAGTAAAAAT 714
QY 175 -----IleGlyValHisGlySerLysLeuSerSerGlyGlyIle--- 187
Db 715 AACGCATGGGTTAATTCCTGTGGCAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 774
QY 188 -----GlyTyrGlnValGly 192
Db 775 ATCTACCGACCGCGGTGGTCAAGAAATACAGGCACATGATGATGATGATGATGATGATG 834
QY 193 GlnSer-----ValAla 196
Db 835 CAAAGTTTGTATAGAGCGGTGGCAACCACTGACCCCAAAATACCGAAACATTTTAAATAG 894
QY 197 GlyValSerGluAsnGlyAsnVal-----LeuAla 206
Db 895 AATGAATGTGCCAATGTAATTATGAGCGGTGTGCTGCTGGCGGTCAAAACCAACTTCAA 954
```

```
QY 207 ArgLeuAspValAspTyrArgThrThrGlyGlyAlaPheAspAlaAsnGly---LysArg 225
Db 955 GCCAAGCCCAACCAATGTCGTGATAGTCAATGTCAAAGATTATACAGGCTCCTAACCCGC 1014
QY 226 IleAlaProGluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsnThr 245
Db 1015 CTTATCCCAAAACCCACTC-----ACCCAAGACAGCAAAATCCTTACTGCTTCGCCCA 1065
QY 246 AsnValAspTrpGlnLeuAspAspLysGlnAsnIleAsn-----LeuAlaLeuThrHis 263
Db 1066 GGT-----TATCAGCTAAACGATGAGCACTATGTCGGTGGTGTATGATAAATACCAAA 1119
QY 264 TyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeu 283
Db 1120 CAAAACCTAGCCCATGCAAGATAAAACCGTGCCTGCTTAT-----CTGGCGGTTTCA 1170
QY 284 PheGlyGluLysProSerLeuAsn-----AlaIleLysGlyLeu 296
Db 1171 GACATTGAAAATCAAGGCTCAGCAACCATGCCCAAGCAATGGCTATATATCAAGGCAAT 1230
QY 297 SerLeuSerGluGlnProLysThrThr-----LysSerThrPheAsnIleAsnTyr 313
Db 1231 AATCTTGGTGAAGCATTCGTGATACCATTTGGGCCAGATTTCAGGTTATGGCATCAACTAT 1290
QY 314 HisHis 315
Db 1291 GCTCAT 1296
RESULT 12
US-08-765-081-4
Sequence 4, Application US/08765081
Patent No. 5798260
GENERAL INFORMATION:
APPLICANT: Tarr, P. I., Bilge, S. S., Besser, T. E., Vary Jr., J. C.
TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
STREET: 2800 Pacific First Center, 1420 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
COMPUTER: IBM PC compatible/Pentium
OPERATING SYSTEM: MS-Windows 3.1
SOFTWARE: Word for Windows-6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08765,081
FILING DATE: March 26, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: June 7, 1995
APPLICATION NUMBER: US 08/265,714
FILING DATE: June 24, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Shelton, Dennis K.
REGISTRATION NUMBER: 26,997
REFERENCE/DOCKET NUMBER: CHOR-1-10286
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-206-682-8100; 1-206-224-0718 (direct)
TELEFAX: 1-206-224-0779
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2091 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
DESCRIPTION: Corresponds to SEQ ID NO:1,
DESCRIPTION: nucleotides 3036-5126
HYPOTHETICAL: NO
```

```

?      ANTI-SENSE: NO
?      ORIGINAL SOURCE:
?      ORGANISM: Escherichia coli O157: H7
?      STRAIN: 86-24 NALR
?      FEATURE:
?      NAME/KEY: CDS
?      LOCATION: 1..2088
US-08-765-081-4

```

Alignment Scores:	
Pred. No.:	9.24e-06
Score:	136.00
Percent Similarity:	37.46%
Best Local Similarity:	21.83%
Query Match:	6.64%
DB:	1
Length:	2092
Matches:	74
Conservative:	53
Mismatches:	118
Indels:	94
Gaps:	18

US-09-889-746-2 (1-400) X US-08-765-081-4 (1-2091)

Qy	40	ValleThrIleAspIysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHis	59
	41		
	42		
	43		
Db	79	GTCGATGATTCCTCGGCATCCGGC-----TATGAGAAAAAGCTGACTAACGACGCCGC	132
	80		
Qy	60	ThrThrIysValIleIeryGluGluGlnIleGlnGluGlnAlaThrGlySerArgGlnLeu	79
	61		
	62		
	63		
Db	133	AGTGTTCGTGATTCAGTCAGAGAGGAATTGCAG-----TCAGCCAGTAC	177
	134		
Qy	80	AlaAspValMetAlaGlnLeu-----IleProSerLeuGlyValSerSergIlyThrThr	97
	81		
	82		
	83		
Db	178	CACGATCTCGCGGAGGCTGTGAGATTCAGTACGATGAGGCTGTGATGTTGAAGAGCTACGGGT	237
	179		
Qy	98	SerAsnProGly-----GlnThrMetHisGlyArgGlnValGlnPheLeu	112
	99		
	100		
Db	238	AAACCAGGAGGCTGGAAATACAGCATCCGAGAGAAATGCGACGACTTACACGCTGATACG	297
	239		
Qy	113	LeuAsnGlyValProLeuThrGlySerArgAspIleSer---ArgGlnLeuAsnSerIle	133
	114		
	115		
Db	298	ATTGATGATGTTTCGTACAGGCGGAGAGCATGACGTGATCCCAACGGTTTTTCTGCGCATG	355
	299		
Qy	132	AsnProAsn-----GlnValAlaArgIleGlnValLeuSergIlyAla	145
	133		
	134		
Db	358	AATACCGGATTCATGCCCCCTCTGGCCGCGCATGTAGAGCTATTGAGGTTATCAGGCGGCGC	411
	359		
Qy	146	ThrSer---IleIeryGlySergIlyAlaThrGlyLeuIleAsnIleValThrIysSer	166
	147		
	148		
Db	418	ATGTCACACCTGTATGGCTCTGATATGCCATGGCGGCTGTGGAAATATCATACACAGAAAG	477
	419		
Qy	165	AspLeuGluGluGlnGlnPheGluThrArgIleGlyValHis-----	178
	166		
	167		
Db	478	AATGACAGCAAAATGGCTCTCTTCGTCATACAGAGGCTGAATTCGACGAAAGCAACA	537
	479		
Qy	179	-----GlySerIysLeuSerSergIlyIle	187
	180		
	181		
Db	538	TGGGTAACAGACAGCAATTAAATTTCGAGACAGTGTCCCTTCGTGATGATTCGTCTC	597
	539		
Qy	188	GlyIryGlnVal-----GlyGlnSerValAlaGlyValSer	199
	189		
	190		
Db	598	AGCCTGCGATACCGCGGTAGACACAACAAGCTGACGGTTCATGCTGCATCATCTAGC	655
	599		
Qy	200	GluAsnGlyAsnValLeuAlaArgLeuAspValAspTyrArgThrThrGlyValAlaPhe	215
	201		
	202		
Db	657	-----	657
	658		
Qy	220	AspAlaAsnGlyIysArgIleAlaProGluProAlaGlnThrAspIysGlnAspSerIys	235
	221		
	222		
Db	658	GATACAGACAGCACCCGTAATTCCTTATCCACAG-----GAGTCACAG	699
	659		
Qy	240	SerLeuSerValAsnThrAsnValAspTyrGlnLeuAspAspIysGlnAsnIleAsnLeu	255
	241		
	242		
Db	700	AATTATATCTTGGTGACAGCTTGACGTGAAGAGCGTGCAGACAGATGTGCTC-----	755
	701		
Qy	260	AlaLeuThrIleTyrAsnAspIysGlnIleAspThrIleAspTyrValAlaProAspTyrGlyAsnArg	275
	261		
	262		

```

Db 754 -----TCGTTTGATGATGATACACAC-----CGGACGCGTTATGATTAACCG 795
Cy 280 LeuAlaValLeuPheGlyGluValProSerLeuAsnAlaIleYsgIleuSerLeuSer 299
Db 796 -----GATGGGGAACCTGGGGAAGTCTTGACCGGGGGAATATGACCGACCTG--- 840
Cy 300 GluGlnProLysThrThrLysSerThrPheAsnIleAsnTrpHisHisAsp----- 316
Db 841 -----CCCTATGAGCGCAAAACAAATTTGAGCTGGCTATATGATCACTTACACCTTC 891
Cy 317 AspLeuTrpGlyAsnThrIleAsnThrAsnAlaTyrTyrArgArgGluYsgIArg 335
Db 892 GGAACATGGAATGCTGATCTGGAATGGAACGAG-----ACAGAAATAAGTCTG 942

RESULT 13
US-09-098-082-4
Sequence 4, Application US/09098082
Patent No. 6040421
GENERAL INFORMATION:
APPLICANT: Tarr, P.I., Blige, S.S., Besser, T.E., Vary Jr., J.C.
TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSER: Christensen, O'Connor, Johnson and Kindness
STREET: 2800 Pacific First Center, 1420 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette-3.5 inch, 1.44MB storage
COMPUTER: IBM PC compatible/Pentium II
OPERATING SYSTEM: MS-Windows 95
SOFTWARE: Word for Windows-6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/098,082
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/765,081
FILING DATE: March 26, 1997
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: June 7, 1995
APPLICATION NUMBER: US 08/265,714
FILING DATE: June 24, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Sheiness, Diana K.
REGISTRATION NUMBER: 35,356
REFERENCE/DOCKET NUMBER: CHOR-1-12402
TELECOMMUNICATION INFORMATION:
TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)
TELEFAX: 1-206-224-0779
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2091 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
DESCRIPTION: Corresponds to SEQ ID NO:1,
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Escherichia coli 0157: H7
STRAIN: 86-24 NMLR
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2088
US-09-098-082-4

Alignment Scores:
Pred. No.: 9.24e-06 Length: 2091
Score: 136.00 Matches: 74

```

Percent Similarity: 37.46% Conservative: 53  
Best Local Similarity: 21.83% Mismatches: 118  
Query Match: 6.64% Indels: 94  
DB: 3 Gaps: 18

US-09-889-746-2 (1-400) x US-09-098-082-4 (1-2091)

Qy 40 ValIleThrIleAspSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHis 59  
Db 79 GTGATGTTCTCGCATCCGC-----TATGAGAAAAGCTGACTAACCGACCGCC 132  
Qy 60 ThrThrLysValIleTyrcLugluGlnIleGlnGlnAlaThrGlySerArgGlnLeu 79  
Db 133 AGTGTTCCTGTTAGTAGCAGGAGAAATGTCAG-----TCCAGCCAGTAC 177  
Qy 80 AlaAspValMetAlaGlnLeu-----IleProSerLeuGlyValSerSerGlyThrThr 97  
Db 178 CAGATCTGGCGAGGCTCTGAGATCAGTAGAGGTGTGATGTTGAAAGTGTAGCGGT 237  
Qy 98 SerAsnPheGly-----GlnThrMetHisGlyArgGlnValGlnPheLeu 112  
Db 238 AAAACCGAGGCTGGAATCAGCATCCGAGGAATGCCAGCCAGTTACACGCTGATACTG 297  
Qy 113 LeuAsnGlyValProLeuThrGlySerArgAspIleSer---ArgGlnLeuAsnSerIle 131  
Db 298 ATTGATGGTTCGTGAGCGGGAAGCAGTACGTCGACTCCCAACGGTTTTTCTGCCATG 357  
Qy 132 AsnProAsn-----GlnValAlaArgIleGluValLeuSerGlyAla 145  
Db 358 ATACCGGTTTCATGCCCTCTGGCGCCATTGAGCGTATTGAGTTATCAGGGGCGG 417  
Qy 146 ThrSer---IleTyrcLysSerGlyAlaThrGlyGlyLeuIleAsnIleValThrLysSer 164  
Db 418 ATGTCACACTGTATGCTCTGATGCGTGGCGGTGTGCTGAATATCATATTACCAAG 477  
Qy 165 AspLeuGluGluGlnPheGluThrArgIleGlyValHis----- 178  
Db 478 AATGCAGACAAATGGCTCTTCCTCGTCAATGCAGGCTGAATCTGCGAAGAACAA 537  
Qy 179 -----GlySerLysLeuSerSerGluGlyIle 187  
Db 538 TGGGGTAACAGCAGCAGTAAATTTCTGGAGCAGTGGTCCCTTGTGATGATCTGTC 597  
Qy 188 GlyTyrcLysVal-----GlyGlnSerValAlaGlyValSer 199  
Db 598 AGCCTGCAGGTACCGGTAGCACACACAGCGTCAGGTTTCATCGCTCATCACTGAGC 657  
Qy 200 GluAsnGlyAsnValLeuAlaArgLeuAspValAspTyrArgThrThrGlyGlyAlaPhe 219  
Db 657 ----- 657  
Qy 220 AspAlaAsnGlyLysArgIleAlaProGluProAlaGlnThrAspLysGlnAspSerLys 239  
Db 658 GATACAGCAGCAGCGGTATT---CTTATCCACG-----GAGTCACAG 699  
Qy 240 SerLeuSerValAsnThrAsnValAspTyrGlnLeuAspAspLysGlnAsnIleAsnLeu 259  
Db 700 AATTATAATCTGGTGCACGCTCTGACTGGAAGCGCTCGAGCAGGATGTGCTC----- 753  
Qy 260 AlaLeuThrHisTyrcLysAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArg 279  
Db 754 -----TGGTTTATATGATGATACACCC-----CGGCAGCGTTATGATACCCGG 795  
Qy 280 LeuAlaValLeuPheGlyGluLysProSerLeuAsnAlaIleLysGlyLeuSerLeuSer 299  
Db 796 -----GATGGCAACTGGGGAGTCTGACCGGGGGATGATACCGGACCGCTG--- 840  
Qy 300 GluGlnProLysThrThrLysSerThrPheAsnIleAsnTyrcHisHisAsp----- 316  
Db 841 -----CGCTATGAGCGAAACAAATTTTCAGCTGGCTGATGATCATCTTTACCTTC 891  
Qy 317 AspLeuTyrGlyAsnThrIleAsnThrAsnAlaTyrcTyrArgArgGluLysGlyArg 335  
Db 941 ----- 941

Db 892 GGAACATGGAATCGTATCTGAACCTGGACGAG-----ACAGAAAATAAAGTCGT 942

## RESULT 14

PCT-US95-06994-4  
; Sequence 4, Application PC/TUS9506994  
; GENERAL INFORMATION:  
; APPLICANT: Children's Hospital & Medical Center  
; APPLICANT: University of Washington  
; APPLICANT: Washington State University Research Foundation  
; APPLICANT: TARR, PHILLIP I  
; APPLICANT: BILGE, SIMA S  
; APPLICANT: BESSER, THOMAS E  
; APPLICANT: VARY JR, JAMES C  
; TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC  
; STREET: SUITE 2800, 1420 FIFTH AVENUE  
; CITY: SEATTLE  
; STATE: WASHINGTON  
; COUNTRY: USA  
; ZIP: WA 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/06994  
; FILING DATE: 07-JUN-95  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/265,714  
; FILING DATE: 24-JUN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BRODERICK, THOMAS F  
; REGISTRATION NUMBER: 31,332  
; REFERENCE/DOCKET NUMBER: CHOR-18591  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (206) 682 8100  
; TELEFAX: (206) 224 0779  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2091 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; DESCRIPTION: corresponds to SEQ ID NO:1,  
; DESCRIPTION: nucleotides 3036-5126  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Escherichia coli O157: H7  
; STRAIN: 86-24 NALR  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..2088  
; PCT-US95-06994-4

Alignment Scores:  
Pred. No.: 9.24e-06 Length: 2091  
Score: 136.00 Matches: 74  
Percent Similarity: 37.46% Conservative: 53  
Best Local Similarity: 21.83% Mismatches: 118  
Query Match: 6.64% Indels: 94  
DB: 5 Gaps: 18

US-09-889-746-2 (1-400) x PCT-US95-06994-4 (1-2091)

Qy 40 ValIleThrIleAspSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHis 59  
Db 79 GTGATGTTCTCGCATCCGC-----TATGAGAAAAGCTGACTAACCGACCGCC 132



```
QY 132 AsnProAsn-----GlnValAlaArgIleGluValLeuSerGlyAla 145
|||
Db 3393 AATACCGGTTTCATGCCCTCTGGCCGCATTCAGCGTATTGAGGTATACGGGGCGG 3452
QY 146 ThrSer---IleTyrGlySerGlyAlaThrGlyLeuIleAsnIleValThrLysSer 164
|||
Db 3453 ATGTCACACTGTATGGCTCTGATCGATGGCGGTGTGGTGAATATCATACCAGAAAG 3512
QY 165 AspLeuGluGluGlnPheGluThrArgIleGlyValHis----- 178
|||
Db 3513 AATGCAGACAAATGGCTCTCTCCGTCATTCAGCGCTGAATCGCAGAAAGCAACAA 3572
QY 179 -----GlySerLysLeuSerSerGluGlyIle 187
|||
Db 3573 TGGGGTAACAGCAGCCAGTTTAAATTTCTGGAGCAGTGGTCCCTTGTGGATGATTCGTG 3632
QY 188 GlyTyrGlnVal-----GlyGlnSerValAlaGlyValSer 199
|||
Db 3633 AGCTGCAGGTACGGGTAGCAGACACACAGCGTCAGCGTTCATCGGTACATCAGTACG 3692
QY 200 GluAsnGlyAsnValLeuAlaArgLeuAspValAspTyrArgThrThrGlyGlyAlaPhe 219
|||
Db 3692 ----- 3692
QY 220 AspAlaAsnGlyLysArgIleAlaProGluProAlaGlnThrAspLysGlnAspSerLys 239
|||
Db 3693 GATACAGCAGCGCGGTATT---CCTTATCCACGC-----GAGTCACAG 3734
QY 240 SerLeuSerValAsnThrAsnValAspTyrGlnLeuAspAspLysGlnAsnLeu 259
|||
Db 3735 AATTATAATCTTGGTCAGCTCTTGACTGGAAGCGTCGAGCAGGATGTCGTC----- 3788
QY 260 AlaLeuThrHisTyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArg 279
|||
Db 3789 -----TGGTTTCATATGGATACACC-----CGGCGCGTTATGATAACCGG 3830
QY 280 LeuAlaValLeuPheGlyGlyLysProSerLeuAsnAlaIleLysGlyLeuSerLeuSer 299
|||
Db 3831 -----GATGGCAACTGGGGAGTCTGACGGGGGATATACCGGACCTGTG--- 3875
QY 300 GluGlnProLysThrThrLysSerThrPheAsnIleAsnTyrHisAsp----- 316
|||
Db 3876 -----CGCTATGAGGAAACAAATAATTCAGCTGGGTATGATCATCTTACCTTC 3926
QY 317 AspLeuTrpGlyAsnThrIleAsnThrAsnAlaTyrTyrArgGluLysGlyArg 335
|||
Db 3927 GGAACATGAAATCGTATCTGAACGTGAACGAG-----ACAGAAATAAAGTGCT 3977

RESULT 16
US-09-098-082-1
; Sequence 1, Application US/09098082
; Patent No. 6040421
; GENERAL INFORMATION:
; APPLICANT: Tarr, P.I., Bilge, S.S., Besser, T.E., Vary Jr., J.C.
; TITLE OF INVENTION: Escherichia Coli 0157:H7 Epithelial Adhesin
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; STREET: 2800 Pacific First Center, 1420 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette-3.5 inch, 1.44Mb storage
; COMPUTER: IBM PC compatible/Pentium II
; OPERATING SYSTEM: MS-Windows 95
; SOFTWARE: Word for Windows-6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,082
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/765,081
```

```
; FILING DATE: March 26, 1997
; APPLICATION NUMBER: PCT/US95/06994
; FILING DATE: June 7, 1995
; APPLICATION NUMBER: US 08/265,714
; FILING DATE: June 24, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: CHOR-1-12402
; TELEPHONE: 1-206-682-8100; 1-206-224-0735 (direct)
; TELEFAX: 1-206-224-0779
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8041 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli O157:H7
; STRAIN: 86-24 NAHR
; IMMEDIATE SOURCE:
; CLONE: PEAR
; US-09-098-082-1

Alignment Scores:
Pred. No.: 8,398-05 Length: 8041
Score: 136.00 Matches: 74
Percent Similarity: 37.46% Conservative: 53
Best Local Similarity: 21.83% Mismatches: 118
Query Match: 6.64% Indels: 94
DB: 3 Gaps: 18

US-09-889-746-2 (1-400) x US-09-098-082-1 (1-8041)
QY 40 ValIleThrIleAspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHis 59
|||
Db 3114 GTGATGATTGTCTCGGCATCCGCG-----TATGAGAAAAGCTGACTAACCCGCCGCC 3167
QY 60 ThrThrLysValIleTyrGluGluGlnIleGlnGluAlaThrGlySerArgGlnIle 79
|||
Db 3168 ACTGTTCTCTGATTAGCAGGAGGAAATTGCAG-----TCCAGCCAGTAC 3212
QY 80 AlaAspValMetAlaGlnLeu-----IleProSerLeuGlyValSerSerGlyThrThr 97
|||
Db 3213 CACGATCTGGCGGAGGCTCTGAGATCAGTAGAGGTGTGGATGTGAAAGTGTACGGGT 3272
QY 98 SerAsnPheGly-----GlnThrMetHisGlyArgGlnValGlnPheLeu 112
|||
Db 3273 AAAACCGGAGGCTCGAAATCAGCATCCGAGAAATCCAGCAGTTACACGCTGATCTG 3332
QY 113 LeuAsnGlyValProLeuThrGlySerArgAspIleSer---ArgGlnLeuAsnSerIle 131
|||
Db 3333 ATTGATGTGTTCGTGTCAGGCGGGAAGAGTACGTGACGTCCCAACGGTTTTTCTGCCATG 3392
QY 132 AsnProAsn-----GlnValAlaArgIleGluValLeuSerGlyAla 145
|||
Db 3393 AATACCGGTTTCATGCCCTCTGGCCGCATTCAGCGTATTGAGGTATACGGGGCGG 3452
QY 146 ThrSer---IleTyrGlySerGlyAlaThrGlyLeuIleAsnIleValThrLysSer 164
|||
Db 3453 ATGTCACACTGTATGGCTCTGATCGATGGCGGTGTGGTGAATATCATACCAGAAAG 3512
QY 165 AspLeuGluGluGlnPheGluThrArgIleGlyValHis----- 178
|||
Db 3513 AATGCAGACAAATGGCTCTCTCCGTCATTCAGCGCTGAATCTGCAGAAAGCAACAA 3572
QY 179 -----GlySerLysLeuSerSerGluGlyIle 187
|||
Db 3573 TGGGGTAACAGCAGCCAGTTTAAATTTCTGGAGCAGTGGTCCCTTGTGGATGATTCGTG 3632
```

```

Qy 188 GlyTyrGlnVal-----GlyGlnSerValAlaGlyValSer 199
      |||||
Db 3633 AGCTGTGAGTACGGCGTACACACAAACGCTCAGGGTTCATCGCTCACATCAGTACG 3692
      |||||
Qy 200 GUAANGlyAsnValLeuAlaArgLeuAspValAspTyrArgThrThrGlyAlaPhe 219
      |||||
Db 3692 ----- 3692
Qy 220 AspAlaAsnGlyLysArgGlyLeaAlaProGluProAlaGlnThrAspLysGlnAspSerLys 239
      |||||
Db 3693 GATACGACGACGACCGCTTT-----CTTATCCACG-----GAGTCACAG 3734
      |||||
Qy 240 SerLeuSerValAsnThrAsnValAspTyrGlnLeuAspLysGlnAsnIleAsnLeu 259
      |||||
Db 3735 AATTAAATCTGTGTCACAGCTTGTGACGAGCGCTCGAGACGAGATGCTC----- 3788
      |||||
Qy 260 AlaLeuThrHisTyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAlaArg 279
      |||||
Db 3789 -----TGCTTTGATATGATGATACACC-----CGGACGCGTTATGATTAACCGG 3830
      |||||
Qy 280 LeuAlaValLeuPheGlyGlyLysProSerLeuAsnAlaIleLysGlyLeuSerLeuSer 299
      |||||
Db 3831 -----GATGGGCACTGGGAGCTGACGGGGGATATGACCGACCTG----- 3875
      |||||
Qy 300 GlnGlnProLysThrThrLysSerThrPheAsnIleAsnTyrHisIleAsp----- 316
      |||||
Db 3876 -----CGCATGACGAGCAACAAATTTCAGCTGGCTATGATCATCTTACACCTTC 3926
      |||||
Qy 317 AspLeuTyrGlyAsnThrIleAsnThrAsnAlaTyrTyrArgArgGlyLysGlyArg 335
      |||||
Db 3927 GGAAACATGAAATCGTATCTGAACGAAACGAG-----ACAGAAATTAAGTCTGT 3977
      |||||
RESULT 17
PCT-US95-06994-1
Sequence 1, Application PC/TUS9506994
GENERAL INFORMATION:
APPLICANT: Children's Hospital & Medical Center
APPLICANT: University of Washington
APPLICANT: Washington State University Research Foundation
APPLICANT: THAR, PHILIP I
APPLICANT: BILGE, SIMA S
APPLICANT: BESSER, THOMAS E
APPLICANT: VARY JR, JAMES C
TITLE OF INVENTION: ESCHERICHIA COLI O157:H7 EPITHELIAL ADHESIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHRISTENSEN, O'CONNOR, JOHNSON AND KINDNESSPLLC
STREET: SUITE 2800, 1420 FIFTH AVENUE
CITY: SEATTLE
STATE: WASHINGTON
COUNTRY: USA
ZIP: WA 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06994
FILING DATE: 07-JUN-95
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/265,714
FILING DATE: 24-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: BRODERICK, THOMAS F
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: CHOR-18591
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682 8100
TELEFAX: (206) 224 0779
INFORMATION FOR SEQ ID NO: 1:

```

```

/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 8041 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ ORIGINAL SOURCE:
/ ORGANISM: Escherichia coli O157:H7
/ STRAIN: 86-24 NAIR
/ IMMEDIATE SOURCE:
/ CLONE: PEAR
/ PCT-US95-06994-1

Alignment Scores:
Pred. No.: 8.39e-05 Length: 8041
Score: 136.00 Matches: 74
Percent Similarity: 37.46% Conservative: 53
Best Local Similarity: 21.83% Mismatches: 118
Query Match: 6.64% Indels: 94
DB: 5 Gaps: 18

US-09-889-746-2 (1-400) x PCT-US95-06994-1 (1-8041)
Qy 40 ValIleThrIleAspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHis 59
      |||||
Db 3114 GTGATGATGTTCTCGGCATCCGC-----TATGAGAAAAGCTCACTTACGACCCGCC 3167
      |||||
Qy 60 ThrThrLysValIleTyrGlnGlnGlnIleGlnGlnAlaThrGlySerArgGlnLeu 79
      |||||
Db 3168 ACTGTTCTGTGATTAAGCAGGAATTCAG-----TCCAGCCAGTAC 3212
      |||||
Qy 80 AlaAspValMetAlaGlnLeu-----IleProSerLeuGlyValSerSerGlyThrThr 97
      |||||
Db 3213 CACGATCTGGCGGAGCGCTGTGATCATGACGAGGTGTGATGTGAAGTGTACGGCT 3272
      |||||
Qy 98 SerAsnPheGly-----GlnThrMetHisGlyArgGlnValGlnPheLeu 112
      |||||
Db 3273 AAAACCGGAGGCGCTGAATATCAGCATCCGAGAAATCCAGCCAGTTACAGCTGATACG 3332
      |||||
Qy 113 LeuAsnGlyValProLeuThrGlySerArgAspIleSer--ArgGlnLeuAsnSerIle 131
      |||||
Db 3333 ATTGATGGTGTTCGTCAGGCGGAGACAGTACGATCCAGCGGTTTCTCCAGT 3392
      |||||
Qy 132 AsnProAsn-----GlnValAlaArgIleGlnValLeuSerGlyAla 145
      |||||
Db 3393 AATACCGGGTTGATGCCCTCTGCGCCGCGCATGAGCGCTATGAGCTTATTCAGGGGCGG 3452
      |||||
Qy 146 ThrSer--IleTyrGlySerGlyAlaThrGlyLysLeuIleAsnIleValThrLysSer 164
      |||||
Db 3453 ATGTCCACACTGATAGTCTTATGCGATGCGCGGTGTGTATATCTTACCAAGAG 3512
      |||||
Qy 165 AspLeuGlnGlnGlnPheGlnThrArgIleGlyValHis----- 178
      |||||
Db 3513 AATGACAGCAAAATGGCTCTTCCGTCATGACGAGCGTGAATTCGACGAAACAAANA 3572
      |||||
Qy 179 -----GlySerLysLeuSerSerGlyLysIle 187
      |||||
Db 3573 TGGGTAACAGACGACGATTAATTCTGAGACAGTGTCCCTTGTGATATCTGTGC 3632
      |||||
Qy 188 GlyTyrGlnVal-----GlyGlnSerValAlaGlyValSer 199
      |||||
Db 3633 AGCTGTGAGTACGGCGTACACACAAACGCTCAGGGTTCATCGCTCACATCAGTACG 3692
      |||||
Qy 200 GUAANGlyAsnValLeuAlaArgLeuAspValAspTyrArgThrThrGlyAlaPhe 219
      |||||
Db 3692 ----- 3692
Qy 220 AspAlaAsnGlyLysArgGlyLeaAlaProGluProAlaGlnThrAspLysGlnAspSerLys 239
      |||||
Db 3693 GATACGACGACGACCGCTTAT-----CTTATCCACG-----GAGTCACAG 3734
      |||||

```

Qy 240 SerLeuSerValAsnThrAsnValAspTyrGlnLeuAspAspLysGlnAsnIleAsnLeu 259  
 Db 3735 AATTATAATCTTGGTGCACGCTCTGACTGGAAGCGCTCGAGCAGGATGTGCTC----- 3788  
 Qy 260 AlaLeuThrHisTyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArg 279  
 Db 3789 -----TGGTTGATATGGATACACC-----CGGCAGGCTTATGATAACCG 3830  
 Qy 280 LeuAlaValLeuPheGlyGlyLysProSerLeuAsnAlaIleLysGlyLeuSerLeuSer 299  
 Db 3831 -----GATGGCAACTGGGGAGCTCTGACGGGGGATATGACCGGACCTG--- 3875  
 Qy 300 GluGlnProLysThrThrLysSerThrPheAsnIleAsnTyrHisHisap----- 316  
 Db 3876 -----CGTATGAGCGGAACAAATTTTCAGCTGGCTATGATCATCATCTTTCACCTTC 3926  
 Qy 317 AspLeuTyrGlyValAsnThrIleAsnThrAsnAlaTyrTyrArgGlyLysGlyArg 335  
 Db 3927 GGAACATGGAATCGTATCTGAACCTGGAACGAG-----ACAGAAAATAAAGTCTGT 3977

## RESULT 18

US-09-453-702B-76/c  
 ; Sequence 76, Application US/09453702B  
 ; Patent No. 6365723  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Blattner, Frederick R.  
 ; Burland, Valerie  
 ; Perna, Nicole T.  
 ; Plunkett, Guy  
 ; Welch, Rod  
 ; TITLE OF INVENTION: No. 6365723el Sequences of E. coli 0157  
 ; NUMBER OF SEQUENCES: 265  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Quatles & Brady  
 ; STREET: 1 South Pinckney Street  
 ; CITY: Madison  
 ; STATE: WI  
 ; COUNTRY: US  
 ; ZIP: 53701-2113

COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.44Mb storage  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Word Perfect 8.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/453,702B  
 ; FILING DATE: 03-Dec-1999  
 ; CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/110,955  
 ; FILING DATE: 04-DEC-1998  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Seay, Nicholas J.  
 ; REGISTRATION NUMBER: 27386  
 ; REFERENCE/DOCKET NUMBER: 960296.95017

TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (608) 251-5000  
 ; TELEFAX: (608) 251-9166

INFORMATION FOR SEQ ID NO: 76:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 6737  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 76:  
 ; US-09-453-702B-76

Alignment Scores:  
 Pred. No.: 8,21e-05 Length: 6737  
 Score: 135.00 Matches: 101  
 Percent Similarity: 36.95% Conservative: 59  
 Best Local Similarity: 23.33% Mismatches: 142

Query Match: 6.60% Indels: 131  
 Db: 4 Gaps: 24  
 US-09-889-746-2 (1-400) x US-09-453-702B-76 (1-6737)  
 Qy 15 LeuSerValAlaValThrGlnGln---LeuTyrAlaGlnProAsnGluSerLeuProThr 33  
 Db 1914 TTATCACTTGGCTTTACCAGCAGCGCTAGCGGCTCAGGAGACGACTCGCTG---ACC 1858  
 Qy 34 ValGluLeuGluProValIleThrIleAspLysSerGlyMetAlaLeuAlaAsnArg 53  
 Db 1857 GTATGGTCAGCTCCGTATCATCGAG----- 1831  
 Qy 54 IleThrGlnMetProHisThrThrLysValIleTyrGluGluGlnIleGlnGluGlnAla 73  
 Db 1830 -----ACGACCACCGTTCTC-----GATCAACCCACCATGAAGGCC 1795  
 Qy 74 ThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSer 93  
 Db 1794 CTGGATAAACAGATGTGCTCAGGCATTAAAGT---GTGTCCTCCCGCGGTGGTGTGCAA 1738  
 Qy 94 SerGlyThrThrSerAsnPheGlyGlnThr-----MetHisGlyArgGlnVal 109  
 Db 1737 AAGTCAGCAGCGCGCAACAGACAGGTAAAGTTCTGGCTTTGATAGTCGGCAGGTG 1678  
 Qy 110 GlnPheLeuAsnGlyValProLeu-----ThrGlySerArgAspIleSer 125  
 Db 1677 CCGGTCTATTTTCGACGGTGTGCCCATTTATGTTCTCCTATGACGGCAACCTCGATCTGGCG 1618  
 Qy 126 ArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyAla 145  
 Db 1617 CGGATCTTGACC-----AACAACTCTGGGGCGAGTCGAAGTTTCCAAAGGGTAT 1570  
 Qy 146 ThrSer---IleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLysSer 164  
 Db 1569 TCGTCGCTGCTTCAGGACCTAATCAGATGGCGGAGCCATTATATCACCACCAGGAAG 1510  
 Qy 165 AspLeuGluGluGlnPheGluThrArgIleGly----- 176  
 Db 1509 -----CCAACAAACCTCTGGAAGCAAGTCTGGGATATCGCAGGGATGGAGCCGTAGC 1456  
 Qy 177 -----ValHisGlySerLysLeuSerSerGluGlyIleGlyTyr--- 189  
 Db 1455 CAGGACATGCTATGATATGATGCTCTTCAATTGTCGCCGAGCAGCGAACTGGGTATTG 1396  
 Qy 190 -----GlnValGlyGlnSerValAlaGlyVal----- 198  
 Db 1395 CAAGTCAGCGGTAGCCAGCTAAAGCAGGATTTTCTCGGCTCCCGCATGGTGAATAAT 1336  
 Qy 199 -----SerGluAsnGlyAsnValLeuAlaArgLeuAspValAspTyrArgThrThr 215  
 Db 1335 GATATTGCAGGCAACACGCGCAAGATGATT----- 1303  
 Qy 216 GlyGlyAlaPheAspAlaAsnGly---LysArgIleAlaProGluProAlaGlnThrAsp 234  
 Db 1302 TCATCGGCTGATGATAAACGCGGCATTGTGAAGCTCGGATTTACACCACGTAACCAACCAT 1243  
 Qy 235 -----LysGlnAspSerLysSerLeuSerVal 243  
 Db 1242 GAATACACACTGACTTACATTAAAGCAGGATGGTGAAGAAAGATAACCCGCCATACAGCGGA 1183  
 Qy 244 AsnThrAsnVal-----AspTyrGlnLeuAspAspLysGlnAsnIleAsn 258  
 Db 1182 AATAGTGGTCAAAAATCAGCTACTCTGCGAGTGGCCAGAGTATGACAAAGAAAGTTTAT 1123  
 Qy 259 Leu---AlaLeuThrHisTyrAsnAspLys-----GlnAsp 269  
 Db 1122 TATCAGGGAACGACCCTAAACCATCGTTTACCTGAAAAGTCGGCTGTATCGCGAC 1063  
 Qy 270 Thr---AspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeuPheGlyGlyLysPro 288  
 Db 1062 ACCTTTGAAAATACCGCTGATGATGTACAACTCGCTGGCTGATTTCGAAAAATAAAGGC 1003



Qy 289 SerLeuAenAlaIlelys-----GlyLeuSerLeuSerGluGlnPro 302  
 Db 1002 ACTGATACGATATATCCGATTACAGCAGCGGTGCGGATTAACACTGCGACGCCGATGTG 943  
 Qy 303 LyrThrThyls---SerThrPheAenIleAenTyHHisIleAenPheLeuTrpGlyAen 321  
 Db 942 CGGAAACGATGCTGCTGCGGTGCGGATTAACACTGGAAGATGAC-----898  
 Qy 322 ThrIleAenThrAenAlaTyTyraArgGluTyGlyArgPheTy+PropheVal-- 340  
 Db 897 -----GRACACCGGAAAGAGTCCGCCGACGCCGCTTACGAT 859  
 Qy 341 -----AlaProPheSerIleAlaIleAlaLeuProIleLe 352  
 Db 858 CGGTATAGATGCTGCTGCTGCTGCGGATTAACACTGATGATGATGATGATGATGAT 799  
 Qy 352 uGlnSerMetAenLeuProSerIleAenAlaIleAenAla 364  
 Db 798 GTCGATGTTGTGGCTGAATGACTATGACTGCGCG 762

RESULT 19

US-08-613-009A-2

; Sequence 2, Application US/08613009A

; Patent No. 6090576

; GENERAL INFORMATION:

; APPLICANT: Myers, Lisa E

; APPLICANT: Schryvers, Anthony B

; APPLICANT: Harkness, Robin E

; APPLICANT: Loosmore, Sheena M.

; APPLICANT: Du, Run-Pan

; APPLICANT: Yang, Yan-Ping

; APPLICANT: Klein, Michel H

; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella

; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sim & McBurney

; STREET: 6th Floor, 330 University Avenue

; CITY: Toronto

; STATE: Ontario

; COUNTRY: Canada

; ZIP: M5G 1R7

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/613,009A

; FILING DATE: 08-MAR-1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Stewart, Michael I

; REGISTRATION NUMBER: 24973

; REFERENCE/DOCKET NUMBER: 1038-542

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (416) 595-1155

; TELEFAX: (416) 595-1163

; INFORMATION FOR SEQ. ID NO. 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3222 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-613-009A-2

Alignment Scores:

Pred. No.: 3,2e-05

Score: 134.00

Percent Similarity: 35.32%

Best Local Similarity: 23.38%

Query Match: 6.55%

Db: 3

Length: 3222

Matches: 94

Conservative: 48

Mismatches: 150

Indels: 110

Gaps: 21

US-09-889-746-2 (1-400) x US-08-613-009A-2 (1-3222)

Qy 10 LeuSerLeuProLeuLeuSerValAla-----ValThrGlnGlnLeu 23  
 Db 64 TGTCTTGGGTCTGCTCTTAACTACAGCAGGTGGCAGCTGGCAACAACAGCCGATAG 123  
 Qy 24 TyrAlaGlnProAenGlnSerLeuProThrValGluLeuGlnProValIleThrIle 43  
 Db 124 GCGAGGCAACAGATTAAGCAAACTTGTGTGCTTGGATGAAGAACTGTGTAAACAGCG 183  
 Qy 44 AspySerGlyMetAlaLeuAlaAenArgIleThrGlnMetProHisThrThrIleVal 63  
 Db 184 AAGAAAAAGCCCGTAA--GCCAAGAAAGTAAAGGCTGTAAAGGTGTCAAAACT 240  
 Qy 64 IleyTglnGlnGlnIleGlnGlnGlnIleThrGlySerArgGlnLeuAlaAenValMet 83  
 Db 241 GCC--GAGACATCAATTAAGAACAGAGCTTAACATTGAGACTA-----285  
 Qy 84 AlaGlnLeuIleProSerLeuGly--ValSerSerGlyThrThrSerAenPheGlyGln 102  
 Db 286 ACACGCTATGACCTCGCATGCTGTGTGAGCAAGGTGCGGGCAAGCTCAGGCTAT 345  
 Qy 103 ThrMetHisGly-----ArgGlnValGlnPheLeuLeuAenGlyVal-----116  
 Db 346 TCTATTGCTGATGATGATTAATAATGCTGTGGGATTTGTTGATGAGCATCAATCAAGCC 405  
 Qy 117 -----ProLeuThrGlySerArg--AspIleSerArgGlnLeu 128  
 Db 406 CAGCACTATGCTTACAGAGCCCTGTGGCAGCAAAATTAATGCGCAGGTGGGCAATC 465  
 Qy 129 AenSerIleAenProAenGlnValAlaArgIleGluValLeuSerGlyVala--ThrSer 147  
 Db 466 AAGCAATGATTAAGCAAAATGCGCTGCTGTGATTAAGTAAAGGTCAAAATCAACT 525  
 Qy 148 IleyTglnSerGlyAlaThrGlyLeuIleAenIleValIleThrIleSer-----164  
 Db 526 GAATAGGCTGTGGGCAATTAATCTGCTGTGGCAATTTGTTACAAACCCGCGATGAC 585  
 Qy 165 -----AspLeuGlnGlnGlnGlnPheGlnThrArg-----174  
 Db 586 ATCATTAAGATGTAAGATGAGGCTGACAGCAAAACCGCTTACGATTAAT 645  
 Qy 175 -----IleGlyValHisGlySerIleLeuSerSerGlyIle--187  
 Db 646 AAGCATGGTTAATCTGTGACAGCAGCAGCAGCAGGTTCTTTAGCGGTCTTATC 705  
 Qy 188 -----GlyTyrGlnValGly 192  
 Db 706 ACTTACACGACCGCGCTGTCAAGATTAACAAGCAGCATGATGATGCTTACAGGGTAC 765  
 Qy 193 GlnSer-----ValAla 196  
 Db 766 CAAGTTTGTATAGAGCGGTGCAACCACTAGCAAAATTAACCAATTTTAAATAGCA 825  
 Qy 197 GlyValSerGlnAenGlyAenVal-----LeuAla 206  
 Db 826 AATGATGCGCAATGATTAATTAAGAGCGGTGCTGCGGTCAAAACCAACTCAA 885  
 Qy 207 ArgLeuAenValAenTyrArgThrThrGlyAlaAenAenAlaAenGly--LysArg 225  
 Db 886 GCCAAGCAACCAATGCGGTGATTAAGTCAATGTCAAAGATTAACAGTCTTAACCCG 945  
 Qy 226 IleAlaProGlnProAlaGlnThrAenPheGlnAenSerIleSerLeuAenThr 245  
 Db 946 CTATCCCAACCACTC-----ACCAAGACAGCAAAATCTTACTGCTCGCCCA 996  
 Qy 246 AenValAenTyrGlnLeuAenAenPheGlnAenIleAen-----LeuAlaLeuThrHis 263  
 Db 997 GGT-----TATCAGTAAACAGTAAAGCACTATGCGGTGTGTGATTAATCAACAAA 1050  
 Qy 264 TyrAenAenPheGlnAenThrAenTyrAlaProAenTyrGlyAenArgLeuAlaValLeu 283  
 Db 1051 CAAACTACGCGCATGCAAGATTAACCGGTGCTGCTTAT-----CTGACGTTTCAAT 1101

Qy 284 PheGlyGluLeuProSerLeuAen-----AlaIleLysGlyLeu 296  
Db 1102 GACATTGAAATCAAGGCTCAGCAACCATGCCAAGCCATGGCTATTATCAAGGCAAT 1161  
Qy 297 SerLeuSerGluGlnProLysThr-----LysSerThrPheAsnIleAenTyr 313  
Db 1162 AATCTTGGTGAAGCATTCGTGATACCATTCAGGTCAGGTCATGAGTATGATGATCAACTAT 1221  
Qy 314 HisHis 315  
Db 1222 GCTCAT 1227  
RESULT 20  
US-08-778-570B-2  
; Sequence 2, Application US/08778570B  
; Patent No. 6437096  
; GENERAL INFORMATION:  
; APPLICANT: Myers, Lisa E  
; APPLICANT: Schryvers, Anthony B  
; APPLICANT: Harkness, Robin E  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Du, Run-Pan  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/778,570B  
; FILING DATE: 03-JAN-1997  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-664  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3222 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-778-570B-2  
Alignment Scores:  
Pred. No.: 3,2e-05 Length: 3222  
Score: 134.00 Matches: 94  
Percent Similarity: 35.3% Conservative: 48  
Best Local Similarity: 23.3% Mismatches: 150  
Query Match: 6.55% Indels: 110  
DB: 4 Gaps: 21  
US-09-889-746-2 (1-400) x US-08-778-570B-2 (1-3222)  
Qy 10 LeuSerLeuProLeuSerValAla-----ValThrGlnGlnLeu 23  
Db 64 TTGTCTTGGTCTGCTTAACATCAGCAGGTGGCAACACACACGCGGATAG 123  
Qy 24 TyrAlaGlnProAsnGluSerLeuProThrValGluLeuGluProValIleThrIle 43

Db 124 GCGGAGGCAACAGATAAGCAAAACCTTGTGTCTGGATGAAACTGTTGTACACGC 183  
Qy 44 AspylSerGlyMetAlaLeuAlaAenArgIleThrGlnMetProHisThrThrLysVal 63  
Db 184 AAGAAACCGCCCGCTAAA--GCCAACGAAGTTACAGGGCTTGGTAAGGTGTCAAACT 240  
Qy 64 IleTyrGluGluGlnIleGlnGluGlnAlaThrGlySerArgGlnLeuAlaAspValMet 83  
Db 241 GCC---GAGACCATCAATAAAGAACAGTCTAAACATTCGAGACTTA----- 285  
Qy 84 AlaGlnLeuIleProSerLeuGly--ValSerSerGlyThrThrSerAsnPheGlyGln 102  
Db 286 ACACGCTATGACCCCTGCGTGTGTGTGAGCAAGCTCGTGGGCAAGTTCAGGCTAT 345  
Qy 103 ThrMetHisGly-----ArgGlnValGlnPheLeuLeuAsnGlyVal----- 116  
Db 346 TCTATTCTGTGTATGATAAAATCGTGTGGCGGTATTGGTTGATGGCATCAATCAAGCC 405  
Qy 117 -----ProLeuThrGlySerArg---AspIleSerArgGlnLeu 128  
Db 406 CAGCACTATGCCCTACAAGGCCCTGTGGCAGGCAAAATTTATGCCAGGTGGGCAATC 465  
Qy 129 AsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyAla---ThrSer 147  
Db 466 AACGAAATAGAAATACGAAATGTCCGCTCGTGTAGATAGTAAAGGTGCAAAATTCAGT 525  
Qy 148 IleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLysSer----- 164  
Db 526 GAATACGCTCTGGGCGCATTTCTGGCTCTGTGGCATTTGTACCAAAACCGCCGATGAC 585  
Qy 165 -----AspLeuGluGluGlnPheGluThrArg----- 174  
Db 586 ATCATCAAGATGTAAGATTTGGGCGCTGCAGACCAAAACCGCCTATGCCAGTAAAAAT 645  
Qy 175 -----IleGlyValHisGlySerLysLeuSerSerGlyGlyIle--- 187  
Db 646 AACGATGGGTAAATTCGTGGCAGCAGCAGCAGGCAAGGAGGTCTTTTGGCGGTATTC 705  
Qy 188 -----GlyTyrGlnValGly 192  
Db 706 ATCTACACGCGCGCTGGTTCAGAATACAAGGCACATGATGATGCTATCAGGGTAGC 765  
Qy 193 GlnSer-----ValAla 196  
Db 766 CAAAGTTTGTAGAGCGGTGGCAACCACTACCCAAATAACCGAACATTTTAAATAGCA 825  
Qy 197 GlyValSerGluAsnGlyAsnVal-----LeuAla 206  
Db 826 AATGAATGTGCAATGGTAATTTATGAGCGGTGTGCTGGCGGTCAAAACCAACTTCAA 885  
Qy 207 ArgLeuAspValAspTyrArgThrThrGlyGlyAlaPheAspAlaAsnGly---LysArg 225  
Db 886 GCCAAGCCCAACCAATGCTGGTGATAGGTCAATGTCAAAGATTATACAGGTCTAACCGC 945  
Qy 226 IleAlaProGluProAlaGlnThrAspLysGlnAspSerLysLeuSerValAenThr 245  
Db 946 CTTATCCCAAAACCCACTC-----ACCCAAGACAGCAAACTCTTACTGCTTCGCCCA 996  
Qy 246 AsnValAspTyrGlnLeuAspAspLysGlnAsnIleAen-----LeuAlaLeuThrHis 263  
Db 997 GGT-----TATCAGCTAAACGATAGCACTATGTCGGTGTGTGTATGAAATCACAAA 1050  
Qy 264 TyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeu 283  
Db 1051 CAAAACTAGCCATCAAGATAAAACCGTGTCTTAT-----CTGACGGTTTCAAT 1101  
Qy 284 PheGlyGluLysProSerLeuAen-----AlaIleLysGlyLeu 296  
Db 1102 GACATTGAAATCAAGGCTCAGCAACCATGCCAAGCCATGGCTATTATCAAGGCAAT 1161  
Qy 297 SerLeuSerGluGlnProLysThr-----LysSerThrPheAsnIleAenTyr 313

Db 1162 AATCTTGGAACGCACTTGCTGATACCATTTGGCCAGATTGAGTATGAGCATCAACTAT 1221  
 Qy 314 H15H15 315  
 Db 1222 GCTCAT 1227

RESULT 21  
 US-09-059-584-2  
 ; Sequence 2, Application US/09059584  
 ; Patent No. 6440701  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Myers, Lisa E  
 ; APPLICANT: Schryvers, Anthony B  
 ; APPLICANT: Harkness, Robin E  
 ; APPLICANT: Loosmore, Sheena M.  
 ; APPLICANT: Du, Run-Pan  
 ; APPLICANT: Yang, Yan-Ping  
 ; APPLICANT: Klein, Michel H  
 ; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
 ; NUMBER OF SEQUENCES: 60  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Sim & McBurney  
 ; STREET: 6th Floor, 330 University Avenue  
 ; CITY: Toronto  
 ; STATE: Ontario  
 ; COUNTRY: Canada  
 ; ZIP: M5G 1R7  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/059,584  
 ; FILING DATE: 14-APR-1998  
 ; CLASSIFICATION:  
 ; PRIORITY APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/778,570  
 ; FILING DATE: 03-JAN-1997  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Stewart, Michael I  
 ; REGISTRATION NUMBER: 24973  
 ; REFERENCE/DOCKET NUMBER: 1038-794  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (416) 595-1155  
 ; TELEFAX: (416) 595-1163  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3222 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; US-09-059-584-2

Alignment Scores:  
 Pred. No.: 3 2e-05 Length: 3222  
 Score: 134.00 Matches: 94  
 Percent Similarity: 35.32% Conservative: 48  
 Best Local Similarity: 23.38% Mismatches: 150  
 Query Match: 6.55% Indels: 110  
 DB: 4 Gaps: 21

US-09-889-746-2 (1-400) x US-09-059-584-2 (1-3222)

Qy 10 LeuSerLeuProLeuLeuSerValAla-----ValThrGlnGlnLeu 23  
 Db 64 TTGTCTTTTGGTCTGCTTAACATCAGCAGTGGCACTGGCAACACAAAGCGCGATAAG 123  
 Qy 24 TYRAlaGlnProAsnGlnSerLeuProThrValGlnLeuGlnProValAlaIleThrIle 43  
 Db 124 GCGAGGCAACAGATTAAGCAAACTTTGTTGTTGCTTGATGATAAACTGTTGTAACAGCG 183

Qy 44 AspIleSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProIleThrThrIleVal 63  
 Db 184 AAGAAAAACGCCCGTAAA---GCCAAGCAAGTTACAGGCGCTGTGAAGTGCTCAAAACT 240  
 Qy 64 IleTyrglnGlnGlnIleGlnGlnAlaThrGlySerArgGlnLeuAlaAspValMet 83  
 Db 241 GCC---GAGACCATCAATTAAGAACAGTGTAAACATTGAGACTTA----- 285  
 Qy 84 AlaGlnLeuIleProSerLeuGly---ValSerSerGlyThrThrSerAsnPhleGln 102  
 Db 286 ACAAGCTATGACCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 345  
 Qy 103 ThrMetHisGly-----ArgGlnValGlnPheLeuLeuAsnGlyVal----- 116  
 Db 346 TCTATTGCTGTATGATTAATAAATGCTGTGGGAGATTGTTGATGAGCATCAATCAAGCC 405  
 Qy 117 -----ProLeuThrGlySerArg---AspIleSerArgGlnLeu 128  
 Db 406 CAGCACTATGCCCTACMAAGCCCTGTGGCAGGCAAAATTATGCGCAGTGGCGCAATC 465  
 Qy 129 AsnSerIleAsnProAsnGlnValAlaArgIleGlnValLeuSerGlyVala---ThrSer 147  
 Db 466 AACGAATATGAATATGAAATATGCTCCCTCCCTGATTAAGTAAAGTCAATTCAGAT 525  
 Qy 148 IleTyrglySerGlyAlaThrGlyLeuIleAsnIleValThrIleSer----- 164  
 Db 526 GAATAGGCTCTGGGGCATTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 585  
 Qy 165 -----AspLeuGlnGlnGlnGlnPheGlnThrArg----- 174  
 Db 586 ATCATCAAAAGATGTAAAGATTGGGCGCTGCAGACCAAAACCGCTATGCCAGTAAAT 645  
 Qy 175 -----IleGlyValHisGlySerIleLeuSerSerGlnGlyIle----- 187  
 Db 646 AACGATGGTTAATTCTGTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 705  
 Qy 188 -----GlyTyrglnValGly 192  
 Db 706 ATCTACACGACCGCGCTGTCAAGAAATACAGGACATGATGATGCTATCAGCTAC 765  
 Qy 193 GlnSer-----ValAla 196  
 Db 766 CAAAGTTTGAATAGACGGTGGCAACACTGACCCAAATATACGAATTTTAATAGCA 825  
 Qy 197 GlyValSerGlnAsnGlyAsnVal-----LeuAla 206  
 Db 826 AATGAATGTCCAAATGTATATTAAGAGCGTGTCTGCTGGCGGTCAACAAACTTCAA 885  
 Qy 207 ArgLeuAspValAspTyrArgThrThrGlyAlaPheAspAlaAsnGly---LysArg 225  
 Db 886 GCCAAGCCAAACCAATGTGTATTAAGTCAATGTCAAAAGATTATACAGTCTCAACCC 945  
 Qy 226 IleAlaProGlnProAlaGlnThrAspIleGlnAspSerLysSerLeuSerValAsnThr 245  
 Db 946 CTTATCCCAACCACTC-----ACCAAGACAGCAATCTTACTGCTGCGCCA 296  
 Qy 246 AsnValAspTyrGlnLeuAspAspLysGlnAsnIleAsn-----LeuAlaLeuThrHis 263  
 Db 997 GGT-----TATCAGCTTAACCAATTAAGCACTATGTGGGTGTGTATGAATACACAAA 1050  
 Qy 264 TYRAsnAspIleGlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeu 283  
 Db 1051 CAAAACTAGCCATGCAATTAACACCGTCTGCTTA-----CTGACGCTTCA 1101  
 Qy 284 PheGlyGlnLysProSerLeuAsn-----AlaIleLysGlyLeu 296  
 Db 1102 GACATTGAAAAATCAAGGCTGACCAACCATGCCAAGCAATGCTATTATCAAGCAAT 1161  
 Qy 297 SerLeuSerGlnGlnProLysThrThr-----LysSerThrPheAsnIleLeuTyr 313  
 Db 1162 AATCTTGGAACGCACTTGCTGATACCATTTGGCCAGATTGAGTATGAGCATCAACTAT 1221  
 Qy 314 H15H15 315

Db 1222 GCTCAT 1227  
|||  
RESULT 22  
US-08-613-009A-1  
; Sequence 1, Application US/08613009A  
; Patent No. 6090576  
; GENERAL INFORMATION:  
; APPLICANT: Myers, Lisa E  
; APPLICANT: Schryvers, Anthony B  
; APPLICANT: Harkness, Robin E  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Du, Run-Pan  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1K7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/613,009A  
; FILING DATE: 08-MAR-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-542  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3438 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-613-009A-1  
Alignment Scores:  
Pred. No.: 3,566-05 Length: 3438  
Score: 134.00 Matches: 94  
Percent Similarity: 35.32% Conservative: 48  
Best Local Similarity: 23.38% Mismatches: 150  
Query Match: 6.55% Indels: 110  
Gaps: 21  
US-09-889-746-2 (1-400) x US-08-613-009A-1 (1-3438)  
QY 10 LeuSerLeuProLeuLeuSerValAla-----ValThrGlnGlnLeu 23  
Db 277 TTGCTTTTGGTCTGCTTAACTACATCAGCAGGTGGCACTGGCAACACACACGCGCGATAAG 336  
QY 24 TyrAlaGlnProAsnGluSerLeuProThrValGluLeuGluProValValIleThrIle 43  
Db 337 GCGGAGGCAACACATAAGACAAACCTTGTGTCTTGGATGAAACTGTTGTAAACAGCG 396  
QY 44 AspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrThrLysVal 63  
Db 397 AAGAAAAACCGCCGCTAAA---GCCAACGAAGTTTACAGGGCTTGGTAAGGTGGTCAAAACT 453  
QY 64 IleTyGluGluGlnIleGlnGluGlnAlaThrGlySerArgGlnLeuAlaAspValMet 83  
Db 454 GCC--GAGACCATCAATAAAGAACAAAGTCTCTAAACATTCGACACTTA-----498

QY 84 AlaGlnLeuIleProSerLeuGly---ValSerSerGlyThrThrSerAsnPhenGlyGln 102  
Db 499 ACACGCTATGACCCCTGGCATTCGTGTGGTTCAGCAAGGTCGTGGGCAAGCTCAGGCTAT 558  
QY 103 ThrMetHisGly-----ArgGlnValGlnPheLeuLeuAsnGlyVal-----116  
Db 559 TCTATTCTGTGTATGGATAAAATCGTGTGGCGGTATTGGTTGATGGCATCAATCAAGCC 618  
QY 117 -----ProLeuThrGlySerArg---AspIleSerArgGlnLeu 128  
Db 619 CAGCACTATGCCCTACAAGGCCCTGTGGCAGGCAAAATATTGCCGAGGTGGGCAATC 678  
QY 129 AsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyAla---ThrSer 147  
Db 679 AACGAATAGAAATACGAAAATGTCCGCTCCGTTCAGATTAGTAAAGGTGCAAAATCAAGT 738  
QY 148 IleTyGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLysSer-----164  
Db 739 GNAATAGGCTCTGGGGCATTTATCTGGCTCTGTGGCATTTGTACCAAAACCGCCGATGAC 798  
QY 165 -----AspLeuGluGluGlnPheGluThrArg-----174  
Db 799 ATCATCAAAGATGTTAAAGATTGGGGCGTGCAGACCAAAACCGCCTATGCCAGTAAAAAT 858  
QY 175 -----IleGlyValHisGlySerLysLeuSerSerGluGlyIle---187  
Db 859 AACGCATGGTTAATTCTGTGGCAGCAGCAGGAGGTTCTTTTAGCGGTCTATC 918  
QY 188 -----GlyTyGlnValGly 192  
Db 919 ATCTACACGCGCGCGTGGTCAAGAAATACAAGGCACATGATGCTATCAGGGTAGC 978  
QY 193 GlnSer-----ValAla 196  
Db 979 CAAAGTTTTGTAGAGCGTGGCAACCACTGACCAAAATAACCGAACATTTTAAATAGCA 1038  
QY 197 GlyValSerGluAsnGlyAsnVal-----LeuAla 206  
Db 1039 AATGAATGTGCCAATGGTAATTATAGGCGGTGTCTGCTGGCGGTCAAAACCAACTTCAA 1098  
QY 207 ArgLeuAspValAspTyArgThrThrGlyGlyAlaPheAspAlaAsnGly---LysArg 225  
Db 1099 GCCAAGCCAAACCAATGTGGTGATAGGTCATGTCAAGATTATACAGGTCTTAAACCGC 1158  
QY 226 IleAlaProGluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsnThr 245  
Db 1159 CTTATCCCAACCCACTC-----ACCCAAGACAGCAAAATCCTTACTGTCTCGCCCA 1209  
QY 246 AsnValAspTrpGlnLeuAspAspLysGlnAsnIleAsn-----LeuAlaLeuThrHis 263  
Db 1210 GGT-----TATCAGCTAAACGATAGCACTATGTCCGTTGTGTATGAATACCAAA 1263  
QY 264 TyrAsnAspLysGlnAspThrAspTyAlaProAspTyArgGlyAsnArgLeuAlaValLeu 283  
Db 1264 CAAAATACGCCATCAAGATAAAACCGTGCCTGCTTAT-----CTGACGGTTCTAT 1314  
QY 284 PheGlyGlyLeuProSerLeuAsn-----AlaIleLysGlyLeu 296  
Db 1315 GACATTGAAAAATCAAGGCTCAGCAACCATGCCCAAGCCAAATGGCTATTATCAAGGCAAT 1374  
QY 297 SerLeuSerGluGlnProLysThrThr-----LysSerThrPheAsnIleAsnTyr 313  
Db 1375 AATCTTGGTGAACGATTCGTGATACCATTTGGCCAGATTTCAGGTTATGGCATCAACTAT 1434  
QY 314 HisHis 315  
Db 1435 GCTCAT 1440

RESULT 23  
US-08-778-570B-1  
; Sequence 1, Application US/08778570B  
; Patent No. 6437096



; APPLICANT: Klein, Michel H  
 ; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
 ; NUMBER OF SEQUENCES: 60  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Sim & McBurney  
 ; STREET: 6th Floor, 330 University Avenue  
 ; CITY: Toronto  
 ; STATE: Ontario  
 ; COUNTRY: Canada  
 ; ZIP: M5G 1R7  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/059,584  
 ; FILING DATE: 14-APR-1998  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/778,570  
 ; FILING DATE: 03-JAN-1997  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Stewart, Michael I  
 ; REGISTRATION NUMBER: 24973  
 ; REFERENCE/DOCKET NUMBER: 1038-794  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (416) 595-1155  
 ; TELEFAX: (416) 595-1163  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3438 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ;

US-09-059-584-1

# Alignment Scores:

Pred. No.: 3,560-05 Length: 3438  
 Score: 134.00 Matches: 94  
 Percent Similarity: 35.32% Conservative: 48  
 Best Local Similarity: 23.38% Mismatches: 150  
 Query Match: 6.55% Indels: 110  
 DB: 4 Gaps: 21

US-09-889-746-2 (1-400) x US-09-059-584-1 (1-3438)

Qy 10 LeuSerLeuProLeuLeuSerValAla-----ValThrGlnGlnLeu 23  
 Db 277 TTGTCTTGGGCTGCTTAACATCAGCGAGGTGGCACTGGCAACACACGCGCGATAAG 336  
 Qy 24 TyrAlaGlnProAsnGluSerLeuProThrValGluLeuGluProValValIleThrIle 43  
 Db 337 GCGGAGGCAACAGATAGACAAACCTTGTGTCTTGATGAAACTGTTGTAAACAGCG 396  
 Qy 44 AspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrThrLysVal 63  
 Db 397 AAGAAAACCGCCGTAAA---GCCACGAAGTTACAGGGCTTGGTAAGGTGGTCAAAACT 453  
 Qy 64 IleTyGluGlnIleGlnGluGlnAlaThrGlySerArgGlnLeuAlaAspValMet 83  
 Db 454 GCC---GAGACCATCAATAAAGAACAAAGTGTCTAAACATTTCGAGACTTA----- 498  
 Qy 84 AlaGlnLeuIleProSerLeuGly---ValSerSerGlyThrThrSerAsnPheGlyGln 102  
 Db 499 ACAGCTATGACCTGGCATGCTGCTGGTTGACCAAGGCTCGTGGGCAAGCTCAGGCTAT 558  
 Qy 103 ThrMetHisGly-----ArgGlnValGlnPheLeuLeuAsnGlyVal----- 116  
 Db 559 TCTATTCTGGTATGATGATAAAATCGTGGCGGTATTGTTGATGCATCAATCAAGCC 618  
 Qy 117 -----ProLeuThrGlySerArg---AspIleSerArgGlnLeu 128

Db 619 CAGCACTATGCCCTACAAGGCCCTGTGGCAGGCAAAAATTATGCGCAGGTGGGCAATC 678  
 Qy 129 AnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyVala---ThrSer 147  
 Db 679 AACGAAATAGAATAACGAAATGTCCGCTCCGTTGAGATTAGTAAAGGTGCAAAATTCAAAT 738  
 Qy 148 IleTyGlySerGlyAlaThrGlyLeuIleAsnIleValThrLysSer----- 164  
 Db 739 GAATACGGCTCTGGGGCATTATCTGGCTCTGTGGCATTGTGTACCAAAACCGCGATGAC 798  
 Qy 165 -----AspLeuGluGluGlnPheGluThrArg----- 174  
 Db 799 ATCATCAAGATGTAAAGATTGGGGCTGCAGACCAAAACCGCTATGCCAGTAAATAAT 858  
 Qy 175 -----IleGlyValHisGlySerLysLeuSerSerGlyGlyIle--- 187  
 Db 859 AACGCATGGGTTAATTCTGTGGCAGCAGCAGGCAAGGAGGTCTCTTTAGCGGTCTTATC 918  
 Qy 188 -----GlyTyGlnValGly 192  
 Db 919 ATCTACACCGCCGCGTGGTCAAGAATACAAGGCACATGATGCTATCAGGGTAGC 978  
 Qy 193 GlnSer-----ValAla 196  
 Db 979 CAAAGTTTGTAGAGCGGTGGCAACCACTACCCAAATAACCGAACATTTTAAATAGCA 1038  
 Qy 197 GlyValSerGluAsnGlyAsnVal-----LeuAla 206  
 Db 1039 AATGAATGTGCCAATGGTAATTATGAGGCGTGTGCTGGCGGTCAAAACCAACTTCAA 1098  
 Qy 207 ArgLeuAspValAspTyrArgThrThrGlyGlyAlaPheAspAlaAsnGly---LysArg 225  
 Db 1099 GCCAAGCCCAACCAATGTGCGTGAAGGTCAATGTCAAGATTTATACAGGTCTTAACCGC 1158  
 Qy 226 IleAlaProGluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsnThr 245  
 Db 1159 CTTATCCCAACCCACTC-----ACCAAGACAGCAAACTCTTACTGCTTCGCCCA 1209  
 Qy 246 AsnValAspTrpGlnLeuAspLysGlnAsnIleAsn-----LeuAlaLeuThrHis 263  
 Db 1210 GGT-----TATCAGCTAAACGATAAGCACTATGTCGGTGGTGTGTATGAAATCACCAAA 1263  
 Qy 264 TyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeu 283  
 Db 1264 CAAAACCTACGCCATCAAGATAAAACCGTGGCTGTCTTAT-----CTGACGGTTTCA 1314  
 Qy 284 PheGlyGluLysProSerLeuAsn-----AlaIleLysGlyLeu 296  
 Db 1315 GACATTGAAAAATCAAGGCTCAGCAACCATGCCCAAGCAATGGCTATTATCAAGGCAAT 1374  
 Qy 297 SerLeuSerGluGlnProLysThr-----LysSerThrPheAsnIleAsnTyr 313  
 Db 1375 AATCTTGTGAACGCAATTCGTGATACCATTTGGGCCAGATTTCAGGTATGGCATCAACTAT 1434  
 Qy 314 HisHis 315  
 Db 1435 GCTCAT 1440

## RESULT 25

; US-08-425-843-6  
 ; Sequence 6, Application US/08425843  
 ; Patent No. 6020154  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Hansen, Eric J.  
 ; APPLICANT: Cope, Leslie D.  
 ; APPLICANT: Jarosik, Gregory P.  
 ; APPLICANT: Hanson, Mark S.  
 ; TITLE OF INVENTION: H. Influenzae Hxub and Hxuc Genes, Proteins  
 ; TITLE OF INVENTION: and Methods of Use  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Arnold, White & Durkee

STREET: P.O. Box 4433  
 CITY: Houston  
 STATE: Texas  
 COUNTRY: USA  
 ZIP: 77210-4433  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/425,843  
 FILING DATE: Concurrently Herewith  
 CLASSIFICATION: 514  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Parker, David L.  
 REGISTRATION NUMBER: 32,165  
 REFERENCE/DOCKET NUMBER: AMCY:012/PAR  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (512) 418-3000  
 TELEFAX: (713) 789-2679  
 TELEX: 79-0924  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 4651 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: DNA  
 US-08-425-843-6

Alignment Scores:  
 Pred. No.: 7.64e-05 Length: 4651  
 Score: 133.00 Matches: 84  
 Percent Similarity: 39.38% Conservative: 55  
 Best Local Similarity: 23.80% Mismatches: 132  
 Query Match: 6.50% Indels: 82  
 Gaps: 18

US-09-889-746-2 (1-400) x US-08-425-843-6 (1-4651)  
 QY 7 pHeGINTpLeuSerleuProleuSeValAlaValThGInGInleuTyraGIn 26  
 Db 733 TTTTCTTAACCTTCCCTCAATACACACCTTAGTGCAGCAATGCACATCGCAA 792  
 QY 27 ProAenGluSerleuProThryValGInleuGInPro-----ValAlleThrIleasp 44  
 Db 793 -----TCCGTTGAATTAGACTTATCAACCTTATTCGACACAGAGAT 834  
 QY 45 lvsSerglMetAlaLeuAlaAsnArglIethrGInMetProHlsthThryValIle 64  
 Db 835 CCAAGTAGTTCCTTATACG----- 855  
 QY 65 TyrglGInGInleGInGInGInAlaThrglSerArgGInleuAlaAspValMetAla 84  
 Db 856 ---CCAGAAAAACAATCTAAGATAGTCTTCTTAAACCAAGCAGCTAGTTCGACGA 912  
 QY 85 GInleu-----IlleProSerleuGlyValSerSerglThrThrSer----- 98  
 Db 913 GCGTTAGAAACACTTCCAAATGTGATGTAGAGGGCGTTGAGAAAGCAATTGCTCAAAA 972  
 QY 99 ---AsnPhelGlyGInThrMetHlsglArgGInValGInPhleuSeuAsnGlyValPro 117  
 Db 973 CCGTAATATCCGAGCGTATAGTATATCGTGTTCGCA---CTCATTTAGCGCGT--- 1026  
 QY 118 LeuThrGlySerArgAspIleSerArgGInleuAsnSerIleAsnPro---AsnGInVal 136  
 Db 1027 ---AGACAAATTTGATTAGACATAGAGGTTCTTATTTCTTCACATGTCATTTATT 1083  
 QY 137 AlaArgIleGInleuSeuSergl---AlaThSerIleTyrglSerGlyAlaThrgl 155  
 Db 1084 CAAGAAATTAAGTATCAAAAGCAACAGTAGAGTCTTATAGGGGTAGCGGTGCTTTGGGT 1143

QY 156 GIlleuIleAsnIleValThrySerAspIleuGInGInGInPhelGInThryArgIle 175  
 Db 1144 GGTGTTGCGCAATGCGTACGCAATAGCTTATGAAATAATGACAAATTC 1203  
 QY 176 GIlValHlsglSerIleuSerSerglGlyIleGlyTyrglGInValGInleuSerVal 195  
 Db 1204 GGAGTT-----AAATTCGCCAA-----GGTTATCAAACTGCTAATATTTTA 1245  
 QY 196 AlaGlyValSerGluAsnGlyAsnValleuAlaArgleuAspValAspTyraThrThr 215  
 Db 1246 TCG-----GAAAGGAGTCTCTGTATTTGCGCAAAATGAC---AAATTCATGTTCTT 1296  
 QY 216 GIlGlyAlaPheAspAlaAsnGlyLysArgIleAlaProGluProAlaGInThryAspLys 235  
 Db 1297 ATTAGGCTTCTTATATATATGCGGTAATTTA-----CGACCTGTTAA 1341  
 QY 236 GInAspSer-----LysSerleuSerValAsnThryAsnValAspTTP 249  
 Db 1342 GCGAACAAAGCTGAATATACACCTTATTAACGTTTGGGGCTTAGCAAAATTTGGTTGG 1401  
 QY 250 GInleuAspAspLysGInAsnIleAsnleuAlaLeuThrHlsthTyraAsnAspLysGInAsp 269  
 Db 1402 CAATTCATGATGCCAACCGTGTGAATTTATCCACCGCAAACTCGTTTAAACAAACA 1461  
 QY 270 ThrAspTyraProAspTyrglYasnArgleuAlaValleuPhelGlyLysProSer 289  
 Db 1462 GCA-----CCAGC 1470  
 QY 290 LeuAsnAlaIleLysGlyLeuSerleuSerglGInProLysThrThrySerThryPhe 309  
 Db 1471 AATAATGAGGTGGAACCAACTTACCAATGAA-----AAATTTATAGAT 1515  
 QY 310 AsnIleAsnTyraHlsthAspAspLeuTyrglIle-----AsnThryIle 323  
 Db 1516 CAATTCATGAGTCTCCAGCGCTCAAAACAGGTTTATCCACAGAGACCAAAACCATATCA 1575  
 QY 324 AsnThryAsnAlaTyraTyraArgGInLysGlyArgPhe 336  
 Db 1576 GAAACATCAGCGTTTACTCTAAGTGAAGAAACACGTTT 1614

RESULT 26  
 US-09-453-702B-174  
 ; Sequence 174, Application US/09453702B  
 ; Patent No. 6365723  
 GENERAL INFORMATION:  
 APPLICANT: Blatner, Frederick R.  
 Burland, Valerie  
 Perna, Nicole T.  
 Plunkett, Guy  
 Welch, Rod  
 TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157  
 NUMBER OF SEQUENCES: 265  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSER: Quarles & Brady  
 STREET: 1 South Pinckney Street  
 CITY: Madison  
 STATE: WI  
 COUNTRY: US  
 ZIP: 53701-2113  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Word Perfect 8.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/453,702B  
 FILING DATE: 03-Dec-1999  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/110,955  
 FILING DATE: 04-DEC-1998  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Seay, Nicholas J.

```

;
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (608) 251-5000
; TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 174:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7304
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 174:
US-09-453-702B-174

```

```

Alignment Scores:
Pred. No.: 0.00313 Length: 7304
Score: 130.50 Matches: 86
Percent Similarity: 36.43% Conservative: 45
Best Local Similarity: 23.96% Mismatches: 133
Query Match: 6.38% Indels: 95
DB: 4 Gaps: 22

```

US-09-889-746-2 (1-400) x US-09-453-702B-174 (1-7304)

```

QY 18 AlaValThrGlnLeuTyrAlaGlnPro-----AsnGluSerLeuProThr 33
DB 5137 GGGCTTTTCTTCTACTGCTTTTGGCACCACCAATGATTCATGCAACCGACACTGCAACGACC 5196
QY 34 ValGluLeuGluProValIleThrIleAspLysSerGlyMetAlaLeuAlaAsnArg 53
DB 5197 AAGATGGCGGAACAATCACTGCTTACAGCGGAT-----GCAAAATACC 5238
QY 54 IleThrGlnMetProHisThrThrLysValIleTyrGluGluGlnIleGlnGlnAla 73
DB 5239 GCAACTGAGCGG-----ACCGATGGTTTATCACTCTGAGC 5274
QY 74 ThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSer----- 89
DB 5275 ACTTCCAGCGGCACATTAACCGATATCCCGATGCTGCGATATCCCGAGGTGGTCAATACG 5334
QY 90 -----LeuGlyValSerSerGlyThr----- 97
DB 5335 GTTAGCGATCAGGTTCTGGAATATCAGATGTCACACCGCTGATGAAACGGCTTTTATAAC 5394
QY 98 ---SerAsnPheGlyGln-----ThrMetHisGlyArgGlnValGlnPheLeuLeuAsn 114
DB 5395 GTCAGTAAGCTGTACAGACCAATACATTAGCGGGAACCTCAGACCGCTTTGTACGTCGT 5454
QY 115 GlyValProLeuThrGlySerArgAsp-----IleSerArgGlnLeuAsnSerIle 131
DB 5455 GGT-----TTTGGTCTAACCGGATGCTCCATCATGACCAACCGCTCGCAACTGTA 5508
QY 132 AsnProAsnGln-----ValAlaArgIleGluValLeuSerGly---AlaThr 146
DB 5509 CTTCTCTCGCAGATTTCACCGCCACAGAACGTTGGAAGTCTAAAGGTCGCGCTCC 5568
QY 147 SerIleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLysSerAspLeu 166
DB 5569 AGCTGTATGGCATTTCTGATCTGCTGATGATTAAGTCGTGACCAAGCGC----- 5622
QY 167 GluGluGluGlnPheGluThrArgIleGlyValHisGlySer---LysLeuSerSerGlu 185
DB 5623 CCGGAAAAAATTC-----CATGGTTCTGTCTCAGCCACGCTCTCC 5664
QY 186 GlyIleGlyTyrGlnValGlyGln---SerValAlaGlyValSerGluAsnGlyAsnVal 204
DB 5665 AGTTTGGCGGAGGCACTGGCGCACTTGATATCAGAGTCCCATTTGAA---GGCACTCAG 5721
QY 205 LeuAlaArgLeuAspValAspTyrArgThrThrGlyAlaPheAspAlaAsn----- 222
DB 5722 CTGGCA-----TACCCTCTCAGCGGGGAAGTCAGGATGAAGATTACTGG 5766

```

```

QY 223 -----GlyLysArgIleAlaProGluProAlaGlnThrAspLysGlnAspSerLys 239
DB 5767 CGAAACTTCGGTAA-----GAGCGC 5787
QY 240 SerLeuSerValAsnThrAsnValAspTyrGlnLeuAspAspLysGlnAsnIleAsnLeu 259
DB 5788 AGTACATTATTGCCCGCTCACTCACTGG---TTTGGTGATTAATGCAACAGTACCATG 5844
QY 260 AlaLeuThrHisTyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArg 279
DB 5845 CTCATTATCCCAT-----CGGACTATATAAACTCCGTTCCGATCGTGA 5886
QY 280 LeuAlaValLeuPheGlyGlyLysProSerLeuAsnAlaIleLysGlyLysSerLeuSer 299
DB 5887 ACGATTTTCGACCTTACGACGAACACAGCCCGCTAAACGTTGATCGAAAAATACGTTTTCAC 5946
QY 300 GluGlnProLysThrThrLys-----SerThrPheAsnIleAsnTyrHis 314
DB 5947 GAAACGTTTAATATACAGATGGTCAGTCCGATCTGGCGCAACTCAACGAGATATCAT 6006
QY 315 HisAspAspLeuTyrGlyAsnThrIleAsnThrAsnAlaTyrTyrArgGluLys 333
DB 6007 CTCATATAGCCAGTGG-----ACAGCGCGCTTTGATTACAGCTACAGCCAGGATAA 6057

RESULT 27
PCT-US96-05320A-264
; Sequence 264, Application PC/TUS9605320A
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences
; APPLICANT: 9410 Key West Avenue
; APPLICANT: Rockville, MD 20850
; APPLICANT: United States of America
; APPLICANT: Johns Hopkins University
; APPLICANT: 720 Rutland Avenue
; APPLICANT: Baltimore, MD 21205
; APPLICANT: United States of America
; APPLICANT: Mark D. Adams
; APPLICANT: Owen White
; APPLICANT: Hamilton O. Smith
; APPLICANT: J. Craig Venter
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome,
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20003-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/05320A
; FILING DATE: April 22, 1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: June 7, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,429
; FILING DATE: June 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Eric K. Steffe
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.014PC01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 264:
; SEQUENCE CHARACTERISTICS:

```



;	LENGTH: 2169 base pairs	
;	TYPE: nucleic acid	
;	STRANDEDNESS: double	
;	TOPOLOGY: linear	
;	PCT-US96-05320A-264	
Alignment Scores:		
Pred. No.:	4.89e-05	Length: 2169
Score:	130.00	Matches: 84
Percent Similarity:	38.79%	Conservative: 51
Best Local Similarity:	24.14%	Mismatches: 137
Query Match:	6.35%	Indels: 76
DB:	5	Gaps: 19
US-09-869-746-2 (1-400) x PCT-US96-05320A-264 (1-2169)		
Qy	7 pHeGInTPLeuSerLeuProLeuLeuSerValAlaValThrgInGInLeuTyAlaGIn	26
Db	7 TTTTCTAAACCTTCCCTTGCATTAACAAACACCTTAGTACAGCAAAATGGCGCTAGGCCAA	66
Qy	27 ProAenGIuSerLeuProThValGIuLeuGIuPro-----ValValIleThrIleAsp	44
Db	67 -----TTCGGTGAATTAACCTCTAACACCTTAATGGACACAGAGAY	108
Qy	45 LysSerGIuMetAlaLeuAlaAsnArgIleThrGInMetProHisThrIleValIle	64
Db	109 CCAMGT-----AGGTGGCTTAATACCCAGAA-----AAACAATCT	144
Qy	65 TyrgInGIuGInIleGInGIuGInAlaThrgIySerArgInLeuAlaAspValMetAla	84
Db	145 AAAGTACTCTCTTCTTAAGCAAGCCACT-----AGTGTTCAGAGTCCCTTGAA	195
Qy	85 GInLeuIleProSerLeuGIuValSerSerGIuThrThrSer-----Asn	99
Db	196 GAC--ATTCCCAATGTTGATGTTAAGAGCGGTTGCAAGACATTCCTCAAAAACCTAAT	252
Qy	100 pHeGInGIuThrMetHisGIuArgGInValGInpHeLeuLeuAsnGIuValProLeuThr	119
Db	253 ATTCGAGGGTTAAGTATGATATATCGTGTGTGCAA--GTCAATGATGGCGTG-----AGA	303
Qy	120 GIySerTAAspIleSerArgGIuLeuAsnSerIleAsnPro--AsnGInValAlaArg	138
Db	304 CAAAATTTTGATTTAGCAATAGAGGTTCTTAATTTCTTCCAATGTCATCATCAACAA	363
Qy	139 IleGIuValLeuSerGIu--AlaThSerIleTyrgIySerGIuAlaThnGIyGIuLeu	157
Db	364 ATTGAAATGATCAAAAGACCAAGTAGCTCTTAATGGGGTACCGGTGACATGGGGTGTT	423
Qy	158 IleAsnIleValThrIySerAspLeuGIuGIuGIuGIuGInpHeGIuThrArgIleGIuVal	177
Db	424 GTGGCAAGCCGTAACCCCAATGCTTTAGACTTAATGAAAAAATAATGACAAATTCGGAGTT	483
Qy	178 HisGIySerIyLeuSerSerGIuGIuIleGIuTyrgInValGIuGInSerValAlaGIy	197
Db	484 -----AAATTCGCCAA-----GATTAATCAA-----	504
Qy	198 ValSerGIuAsnGIuAsnValIleuAlaArgLeuAspValaAspTyArgThrThrgIyGIy	217
Db	505 -----ACTGCAATATTAATTAATCGAAAAAGAGATGTTCTGATTTGGCGCAAAAGAC	555
Qy	218 AlaIleAspAla-----AsnGIuYlyAsnArgIleAlaProGIuProAlaGInThrAspLys	235
Db	556 AAATTCGATGTCTTATATAGTGTTCTTAATATATATAGCGAGATTTTAAACGACGTGTAA	615
Qy	236 GInAspSer-----LysSerLeuSerValaAsnThrAsnValaAspTrp	249
Db	616 GGCAACAAAGCTAAATATACCGCTATATACAGTTTGGGGGCTTAACAAATTCGGTTGG	675
Qy	250 GInLeuAspAspLysGIuAlaAsnIleAsnIleuAlaLeuThrHisTyArgAsnAspLysGIuAsp	268
Db	676 CAATATTAATGATGCAACCGCGTGGAATTAATCCACCGCGCAAACTCGTTTAAACAAACA	735
Qy	270 ThrAspTyAlaIleProAspTyGIuAsnArgLeuAlaValaLeuPheGIuGIuLysProSer	289

```
D8      736 GCA-----CGAGC 744||| ||| |
Qy     290 LeuSenAlaiLeysGlyLueSerLeusErGlInProLySerrThrTyLSerThrPhe 3090
Db     745 AATATGTGGTGCCTTAACACTCATGAATGACA-----ATTACAGATCAAATCACAA 798
Qy    310 AsnIleAsnTrHShIsEApSplEuTriPolYasnThrIleasnthrrAenAlatYtyr 3229
Db   799 AAGTTCCAGGACAAAAGAAGATCTTCCTCCTAACACAAACATCATCCA-----852
Qy    330 ArgArgGluYSglyArgSPhetyr 337
Db   853 ---TCAGAAAGATCAGAGTTTTAC 873

RESULT 28
US-08-613-009A-6
; Sequence 6, Application US/08613009A
; Patent No. 6090576
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Harkness, Robin E
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Du, Run-Pan
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
ADDRESS: Sim & Mcburney
STREET: 6th Floor, 330 University Avenue
City: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/613,009A
FILING DATE: 08-MAR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24973
REFERENCE/DOCKET NUMBER: 1038-542
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3210 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-613-009A-6

Alignment Scores:
Pred. NO.:          9,29e-05           Length:        3210
Score:             130.00              Matches:         93
Percent Similarity: 35.07%            Conservative:    48
Best Local Similarity: 23.13%         Mismatches:     151
Query Match:       6.35%               Indels:         110
DB:                3                  Gaps:           21

US-08-889-746-2 (1-400) x US-08-613-009A-6 (1-3210)
Qy     10 LueSrLeuPrOleuSeValAla-----ValThrGInglnLeu 23
D8     55 TTGCTTTTGSGGTGCTTAAATACAGCAGGTGSCACTGCCAAACACACGCGCGATPAg 114
```

Qy 24 TyrAlaGlnProAsnGluSerLeuProThrValGluLeuGluProValValleThrle 43  
Db 115 GCGGAGGCAACAGTAAGACAAACCTTGTGTCTTGGATGAAACTGTGTAAACAGCG 174  
Qy 44 AspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrThrLysVal 63  
Db 175 AAGAAAAACGCGCGTAA--CCACAGAGTTACAGGGCTTGGTAAGGTGTCACAAACT 231  
Qy 64 IleTyrGluGluGlnIleGlnGlnAlaThrGlySerArgGlnLeuAlaAspValMet 83  
Db 232 GCC--GAGACATCAATAGACAGAGCTTAACATTCGAGACTTA----- 276  
Qy 84 AlaGlnLeuIleProSerLeuGly---ValSerSerGlyThrThrSerAsnPheGlyGln 102  
Db 277 ACACGCTATGACCTGGCATTCGTGGTTGAGCAAGGTCGTGGGCAAGCTCAGGCTAT 336  
Qy 103 ThrMetHisGly-----ArgGlnValGlnPheLeuLeuAsnGlyVal----- 116  
Db 337 TCTATTCTGTGTATGGATAAAATCGTGGCGGTATGTTGATGGCATCAATCAAGCC 396  
Qy 117 -----ProLeuThrGlySerArg---AspIleSerArgGlnLeu 128  
Db 397 CAGCACTATCCCTACAGGCTCTGGGAGGCAAAATATATCCCGCAGGTGGGCAATC 456  
Qy 129 AsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyAla---ThrSer 147  
Db 457 AACGAAATAGAAATACGAAATGTCGCTCGCTGAGATTAGTAAGGTGCAAAATTCAGT 516  
Qy 148 IleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLysSer----- 164  
Db 517 GAATACGGCTCTGGGCAATATCTGCTGCTGGCATTTGTTACCAAAACCCGCGATGAC 576  
Qy 165 -----AspLeuGluGluGlnPheGluThrArg----- 174  
Db 577 ATCATCAAGATGTTAAGATTTGGGCGGTGCAGACCAAAACCGCTATCCAGTAAAT 636  
Qy 175 -----IleGlyValHisGlySerLysLeuSerSerGluGlyLe----- 187  
Db 637 AACGATGGTTAATTCCTGTGGCAGCAGCAGGAGGAGGTTCTTTTAGCGGTCTATC 696  
Qy 188 -----GlyTyrGlnValGly 192  
Db 697 ATCTACCGACCGCGTGTCAAGAAATCAAGGCACATGATGATGCCTATCAGGATGAC 756  
Qy 193 GlnSer-----ValAla 196  
Db 757 CAAGTTTGTATAGACGGTGGCAACCACTGACCCAAATACCCAAATTTTAAATAGCA 816  
Qy 197 GlyValSerGluAsnGlyAsnVal-----LeuAla 206  
Db 817 AATGAATGTGCAATGGTATATGAGCGGTGTGCTGGCGGTCAAAACCAAACTCCAA 876  
Qy 207 ArgLeuAspValAspTyrArgThrThrGlyGlyAlaPheAspAlaAsnGly---LysArg 225  
Db 877 GCTAAGCCCAACCAATGTGGTGATAGAGTCAATGTCAAAAGATTATACAGGTCCTAACCC 936  
Qy 226 IleAlaProGluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsnThr 245  
Db 937 CTTATCCCAACCCACTC-----ACCAAGACAGCAAAATCTTACTGCTGCGCCA 987  
Qy 246 AsnValAspTrpGlnLeuAspAspLysGlnAsnIleAsn-----LeuAlaLeuThrHis 263  
Db 988 GGT-----TATCAGCTAAACGATAGCACTATGTCGGTGGTGTGTGTAATCAACAAA 1041  
Qy 264 TyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeu 283  
Db 1042 CAAAACCTACGCCCATGCAAGATAAAACCGCTGCTGTAT-----CTGACGGTTCAT 1092  
Qy 284 PheGlyGluLysProSerLeuAsn-----AlaIleLysGlyLeu 296  
Db 1093 GACATTGAAAATCAAGGCTCAGCAACCATGCGCAAGCCCAATGGCTATTATCAAGCAAT 1152

Qy 297 SerLeuSerGluGlnProLysThr-----ThrLysSerThrPheAsnIleAsnTyr 313  
Db 1153 AACCTTGTGTGAACGCAATTCGTGATGCCATCGGCAAAATTCAGGTTATGGCATCAACTAT 1212  
Qy 314 HisHis 315  
Db 1213 GCTCAT 1218  
RESULT 29  
US-08-778-570B-6  
; Sequence 6, Application US/08778570B  
; Patent No. 6437096  
; GENERAL INFORMATION:  
; APPLICANT: Myers, Lisa E  
; APPLICANT: Schryvers, Anthony B  
; APPLICANT: Harkness, Robin E  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Du, Run-Pan  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/778,570B  
; FILING DATE: 03-JAN-1997  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-664  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3210 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-778-570B-6  
Alignment Scores:  
Pred. No.: 9.29e-05 Length: 3210  
Score: 130.00 Matches: 93  
Percent Similarity: 35.07% Conservative: 48  
Best Local Similarity: 23.13% Mismatches: 151  
Query Match: 6.35% Indels: 110  
Gaps: 21  
US-09-889-746-2 (1-400) x US-08-778-570B-6 (1-3210)  
Qy 10 LeuSerLeuProLeuLeuSerValAla-----ValThrGlnGlnLeu 23  
Db 55 TTGCTTTTGGTCTCTTAAACATCAGCAGGTCGGCAACACCAACGCGCCGATAAG 114  
Qy 24 TyrAlaGlnProAsnGluSerLeuProThrValGluLeuGluProValValleThrle 43  
Db 115 GCGGAGGCAACAGTAAGACAAACCTTGTGTGCTTGGATGAAACTGTGTAAACAGCG 174  
Qy 44 AspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrThrLysVal 63

```

Db 175 MAGAAAAAGCCCGTAAA---GCCAAGAAATTACAGGCGTGTGTAAGTGTGCAAAACT 231
Qy 11eTyrglUglnlInleglnUglnAlaThrglySerArglnleuAlaAspValMet 83
Db 232 GCC---GAGGCAATCAATAAGAAAGAGTCTAAACATTGAGACTTA----- 276
Qy 84 AlaGlnleuIleProSerleuGly---ValSerSerGlyThrThrSeranpheGlyGln 102
Db 277 ACAAGCTATGACCTGCTGATTCGTGTGAGCAAGGTGTGGGGCAAGCTCAGGCTAT 336
Qy 103 ThMeHiGly-----ArgGlnValGlnPheleuLeuAnglyVal----- 116
Db 337 TCAATTCGTGTGATGATAAAAATCGTGTGGCGTATTTGTTATGATGCAATCAAGCC 396
Qy 117 -----ProleuThrglySerArg---AspIleSerArgGlnLeu 128
Db 397 CAGCACTATGCCCTACAGAGCCCTGTGTGCGCAAGAAATTAATGCCGAGGTGGGCAATC 456
Qy 129 AasnSerIleasnProasnGlnValAlaArgIleGlnValleuSerGlyAla---ThrSer 147
Db 457 AACGAATAGAAATACGAAAAATGTCGCTCGTTGAGATTAAGTGAAGTGCAGAAATTCAGT 516
Qy 148 11eTyrglySerGlyAlaThrglyGlyLeuIleasnIleValThrlySer----- 164
Db 517 GAATACGGCTCTGGGCAATTATCTGCTGTGTGCAATTTTGTACCAAAACCGCGCATGAC 576
Qy 165 -----AspleuGlnUglnUglnPheGlnThrglyArg----- 174
Db 577 ATCATCAAGATGTAAAGTTTGGGGCTGTGCAAGCAAAACCCGCTATGCGCACTAAAAAT 636
Qy 175 -----11eGlyValH1eGlySerlySerleuSerGlnGlyIle--- 187
Db 637 AACGCATGGGTATATTCTGTGCGAGCAGCAGCAAGGCAAGGCTTTTACGGGCTTATC 696
Qy 188 -----GlyTyrglnValGly 192
Db 697 ATTACACCGACCGCGGTGTCAAGAAATACAGCAGCATGATGATGCTTACAGGGTAGC 756
Qy 193 GlnSer-----ValAla 196
Db 757 CAAGTTTGTATAGACGGGTGCAACACCATGACCCCAATTAACCAAAATTTTAAATAGCA 816
Qy 197 GlyValSerGlnAnglyAsnVal-----LeuAla 206
Db 817 AATGAATGTGCCATGTATATTATGAGGCGTGTGCTGTGGCGTCAACCAAACTCCAA 876
Qy 207 ArgleuAspValAspTyrrArgThrThrGlyAlaPheAspAlaAngly---LysArg 225
Db 877 GCTAAGCCCAACCAATGTGCTGTATAGTCAATGTCAAAAGATTATACAGGTCTTAACCGC 936
Qy 226 11eAlaProGlnProAlaGlnThrAspLyGlnAspSerlySerleuSerValAsnThr 245
Db 937 CTATATCCCAACCACTC-----ACCCAAGACAGCAAACTCTTACTGCTTCCGCCA 967
Qy 246 AsnValAspTyrrGlnleuAspAspLyGlnAsnIleAsn-----LeuAlaLeuThrHis 263
Db 988 GGT-----TATCAGCTAAACGATTAAGCACTATGTGCGTGTGTATGAATACACCAAA 1041
Qy 264 TyrrAsnAspLyGlnAspThrAspTyrrAlaProAspTyrrGlyAsnArgleuAlaValLeu 283
Db 1042 CAATAATAGCCCATGCAAGATTAATAACCGTGTCTGCTTAT-----CTGACGGTTCAT 1092
Qy 284 PheGlyGlnlyProSerleuAsn-----AlaIlelySerGlyLeu 296
Db 1093 GACATTGAAATCAAGGCTCAGCAACCATGCGCAAGCCCATGCTATTAACAAGCAAT 1152
Qy 297 SerleuSerGlnProlyThr-----ThrlySerThrPheAsnIleAsnTyrr 313
Db 1153 AACCTTGGTGAACGCAATTCGTATGTCATTTGGGGCAAAATTCAGGTTATGCACTCACTAT 1212
Qy 314 H1eHis 315
Db 1213 GCTCAT 1218

```

```

RESULT 30
US-09-059-584-6
Sequence 6, Application US/09059584
Patent No. 6440701
GENERAL INFORMATION:
APPLICANT: Myers, Lisa E
APPLICANT: Schryvers, Anthony B
APPLICANT: Harkness, Robin E
APPLICANT: Loosmore, Sheena M.
APPLICANT: Du, Run-Pan
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/059,584
FILING DATE: 14-APR-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/778,570
FILING DATE: 03-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24973
REFERENCE/DOCKET NUMBER: 1038-794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 3210 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-059-584-6
Alignment Scores:
Pred. No.: 9,29e-05 Length: 3210
Score: 130.00 Matches: 93
Percent Similarity: 35.07% Conservative: 48
Best Local Similarity: 23.13% Mismatches: 151
Query Match: 6.35% Indels: 110
Gaps: 21
US-09-889-746-2 (1-400) x US-09-059-584-6 (1-3210)
Qy 10 LeuSerleuProleuLeuSerValAla-----ValThrGlnGlnLeu 23
Db 55 TTCTCTTTGGTCTGCTTAACTACGCGAGGTGCGCACTGCGCAACCAACGCGCGCATAG 114
Qy 24 TyrrAlaGlnProasnGlnleuSerleuProThrValGlnleuGlnProValIleThrIle 43
Db 115 GCGAGAGCAAGATTAAGCAAACTTGTGTGTGCTTGATGAACCTGTTGAACAGCG 174
Qy 44 AspLySerGlyMeAlaLeuAlaAsnArgIleThrglnMeCProHisThrThrlySerVal 63
Db 175 MAGAAAAAGCCCGTAAA---GCCAAGAAATTACAGGCGTGTGTAAGTGTGCAAAACT 231
Qy 64 11eTyrglUglnlInleglnUglnAlaThrglySerArglnleuAlaAspValMet 83

```

```
Db 232 GCC---GAGACCATCAATAAGAAACAGGTCTAAACATTCGAGACTTA----- 276
Qy 84 AlaGlnLeuLeuProSerLeuGly---ValSerSerGlyThrThrSerAsnPheGlyGln 102
Db 277 ACACGCTATGACCTGGCATTCTGCTGTTGAGCAAGGTCTGGGGCAAGCTCAGGCTAT 336
Qy 103 ThrMetHisGly-----ArgGlnValGlnPheLeuLeuAsnGlyVal----- 116
Db 337 TCTATTCTGGTGTGATAAATAATCGTGGCGGTATTGTTGATGGCATCAATCAAGCC 396
Qy 117 -----ProLeuThrGlySerArg---AspIleSerArgGlnLeu 128
Db 397 CAGCACTATGCCCTACAAGCCCTGTGGCAGCAAAAATATATGCCGAGGTGGGCAATC 456
Qy 129 AsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyVala---ThrSer 147
Db 457 AACGAATAGATACGAATAATGTCGCTCCGCTGAGATTAGTAAGGTGCAAAATTCAGT 516
Qy 148 IleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLysSer----- 164
Db 517 GAATACGGCTCTGGGCAATTATCTGCTCTGTGGCATTTGTTACCAAAACCCGCGATGAC 576
Qy 165 -----AspLeuGluGluGlnPheGluThrArg----- 174
Db 577 ATCATCAAGATGGTAAGATGGGGCGGTGCAGACCAAAACCCCTATGCCAGTAAAT 636
Qy 175 -----IleGlyValHisGlySerLysLeuSerSerGlyGlyIle--- 187
Db 637 AACGATGGTTAATCTGTGCGCAGCAGCAGCAGGAGGTCTTTAGCGGTCTATC 696
Qy 188 -----GlyTyrGlnValGly 192
Db 697 ATCTACCGACCGCGTGTGTCGAAGTAACAAGGCACATGATGATGCTTATCAGGCTAGC 756
Qy 193 GlnSer-----ValAla 196
Db 757 CAAGTTTGTATGAGCGGTGGCAACCACTGACCCAAATAACCCAAATTTTAAATAGCA 816
Qy 197 GlyValSerGluAsnGlyAsnVal-----LeuAla 206
Db 817 AATGATGTCCATGTTAATATGAGCGGTGTGCTGCTGGCGGTCAAAACCACTCCAA 876
Qy 207 ArgLeuAspValAspTyrArgThrThrGlyGlyAlaPheAspAlaAsnGly---LysArg 225
Db 877 GCTAAGCCAAACCAATGTGGTGATAGGTCAATGTCAAGATTTATACAGGTCTTAACCGC 936
Qy 226 IleAlaProGluProAlaGlnThrAspLysGlnAspSerLysSerLysSerValAsnThr 245
Db 937 CTTATCCCAACCCACTC-----ACCCAAGCAGCAAAATCTTCTGCTTCGCCCA 987
Qy 246 AsnValAspTrpGlnLeuAspAspLysGlnAsnIleAsn-----LeuAlaLeuThrHis 263
Db 988 GGT-----TATCAGTTAAACGNTAAGCACTATGTCGGTGGTGTGATGAATCACCANA 1041
Qy 264 TyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeu 283
Db 1042 CAAAACCTACGCCATGCAAGATAAACCCTGCTGCTTAT-----CTGACGGTTCAT 1092
Qy 284 PheGlyGluLysProSerLeuAsn-----AlaIleLysGlyLeu 296
Db 1093 GACATTGAAAAATCAAGGCTCAGCAACCACTATGGCCAGCAACCAATGGCTATTATCAAGGCAAT 1152
Qy 297 SerLeuSerGluGlnProLysThr-----ThrLysSerThrPheAsnIleAsnTyr 313
Db 1153 AACCTTGGTGAACGCATTCTGTGATGCCATTGGGCAAAATTCAGTTATGGCATCAACTAT 1212
Qy 314 HisHis 315
Db 1213 GCTCAT 1218
RESULT 31
US-08-613-009A-5
```

```
; Sequence 5, Application US/08613009A
; Patent No. 6090576
; GENERAL INFORMATION:
; APPLICANT: Myers, Lisa E
; APPLICANT: Schryvers, Anthony B
; APPLICANT: Harkness, Robin E
; APPLICANT: Loosmore, Sheena M.
; APPLICANT: Du, Run-Pan
; APPLICANT: Yang, Yan-Ping
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: 6th Floor, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/613.009A
; APPLICATION NUMBER: US/08/613.009A
; FILING DATE: 08-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24973
; REFERENCE/DOCKET NUMBER: 1038-542
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3660 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-613-009A-5
```

```
Alignment Scores:
Pred. No.: 0.000115 Length: 3660
Score: 130.00 Matches: 93
Percent Similarity: 35.07% Conservative: 48
Best Local Similarity: 23.13% Mismatches: 151
Query Match: 5.35% Indels: 110
DB: 3 Gaps: 21
```

```
US-09-889-746-2 (1-400) x US-08-613-009A-5 (1-3660)
Qy 10 LeuSerLeuProLeuLeuSerValAla-----ValThrGlnGlnLeu 23
Db 322 TTGCTTTGGTCTCTTAAACATCACGAGGTGGCAACACACACGCCGCAATAG 381
Qy 24 TyrAlaGlnProAsnGluSerLeuProThrValGluLeuGluProValIleThrIle 43
Db 382 CGCGAGGCAACAGATAAGACAAACCTTGTGTGCTTGGATGAAACCTGTTCTAACAGCG 441
Qy 44 AspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrThrLysVal 63
Db 442 AAGAAAAACCCCGCTAAA---GCCAACCAAGTTACAGGCTTGGTAAGGTGTCACAAACT 498
Qy 64 IleTyrGluGluGlnIleGlnGlnAlaThrGlySerArgGlnLeuAlaAspValMet 83
Db 499 GCC---GAGACCATCAATAAAGAACAGTGTAAACATTCGAGACTTA----- 543
Qy 84 AlaGlnLeuIleProSerLeuGly---ValSerSerGlyThrThrSerAsnPheGlyGln 102
Db 544 ACACGCTATGACCTGGCATTCGTGTTGAGCAAGGTCTGTTGGGGCAAGCTCAGGCTAT 603
```

103 ThrMetHisGly-----ArgGlnValGlnPheLeuLeuAsnGlyVal----- 116  
 Db 604 TCTATTGCGTATGATGATTAATAATCGTGTGGCGGTATGCTGATGATGATCAATCAAGCC 663  
 Qy 117 -----ProLeuThrGlySerArg---AspIleSerArgGlnLeu 128  
 Db 664 CAGCACTATGCCCTTACAGAGCCCTGTGGGAGCAAAATTATGCGGAGGTGGGCAATC 723  
 Qy 123 AasnSerIleAsnProAsnGlnValAlaArgIleGlnValLeuSerGlyVala---ThrSer 147  
 Db 724 AACGAAATAGAAATAGCAAAATGTCGCCCTCCGTTGAGATTAAGTAAAGTGAATTCAGCT 783  
 Qy 148 ILeTyrGlySerGlyAlaThrGlyLeuIleAsnIleValThrIleSer----- 164  
 Db 784 GAATACGGCTCTGGGCGCATTAATCTGCTGTGTCGATTTGTTACAAACCCGCGATGAC 843  
 Qy 165 -----AspLeuGlnGluGlnPheGlnThrArg----- 174  
 Db 844 ATCATCAAAAGATGTAAGATTTGGGCGGTGCAAGACCAAAACCGCTATGCGCAATAAAT 903  
 Qy 175 -----IleGlyValHisGlySerIleSerSerGlyGlyIle--- 187  
 Db 904 AAGCATGGGTTAATCTCTGTGGAGAGAGAGCAAGGAGGTTCTTTAGCGGCTTTATC 963  
 Qy 188 -----GlyTyrGlnValGly 192  
 Db 964 ATCTACACCGACCGCGGTGTCAGAAATACAAAGACATGATGATGCTATGCGGTAGC 1023  
 Qy 193 GlnSer-----ValAla 196  
 Db 1024 CAAAGTTTGTATAGAGCGGTGCAACCACTGACCAAAATTAACCAAAATTTTAAATAGCA 1083  
 Qy 197 GlyValSerGlnAsnGlyAsnVal-----LeuAla 206  
 Db 1084 AATGATGTGCGCATGTATTAATTAAGCGGTGCTGCGGTGCAACCAACTCCAA 1143  
 Qy 207 ArgLeuAspValAspTyrArgThrThrGlyAlaPheAspAlaAsnGly---LysArg 225  
 Db 1144 GCTAACGCCAACATGTGCTGATAGTCAATGTCAAGTATATACAGTCTTAACCGC 1203  
 Qy 226 ILeAlaProGlnProAlaGlnThrAspLysGlnAspSerIleSerSerValAsnThr 245  
 Db 1204 CTATCCCAAGCCCACTC-----ACCCAAAGACAGCAAAATCTTACTGCTTGGCCCA 1254  
 Qy 246 AasnValAspTyrGlnLeuAspLysGlnAsnIleAsn-----LeuAlaLeuThrHis 263  
 Db 1255 GGT-----TATCAGCTTAAACATTAAGCACTATGTCGGTGTGTATGAAATCACCAAA 1308  
 Qy 264 TyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeu 283  
 Db 1309 CAAATCTAGCGCATGCAAGATTAACCGTCTGCTTAT-----CTGACGGTTTCA 1359  
 Qy 284 PheGlyGlnLysProSerLeuAsn-----AlaIleLysGlyLeu 296  
 Db 1360 GACATTGAAATAACAAAGCTCAGCAACATGCGCAAGCCAAATGCTATTATCAAGCAAT 1419  
 Qy 297 SerLeuSerGlnGlnProLysThr-----ThrLysSerThrPheAsnIleAsnTyr 313  
 Db 1420 AACCTTGGTGAAGCATCTGTATGCTATGCGGCAAAATTCAGATTATGCGATCACTAT 1479  
 Qy 314 HisHis 315  
 Db 1480 GCTCAT 1485

APPLICANT: Du, Run-Pan  
 APPLICANT: Yang, Van-Ping  
 APPLICANT: Klein, Michel H  
 TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
 NUMBER OF SEQUENCES: 43  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: 6th Floor, 330 University Avenue  
 CITY: Toronto  
 STATE: Ontario  
 COUNTRY: Canada  
 ZIP: M5G 1R7  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/778,570B  
 FILING DATE: 03-JAN-1997  
 CLASSIFICATION: 536  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Stewart, Michael I  
 REGISTRATION NUMBER: 24973  
 REFERENCE/DOCKET NUMBER: 1038-664  
 TELEPHONE: (416) 595-1155  
 TELEFAX: (416) 595-1163  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 3660 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 US-08-778-570B-5  
 Alignment Scores:  
 Pred. No.: 0.000115 Length: 3660  
 Score: 130.00 Matches: 93  
 Percent Similarity: 35.07% Conservative: 48  
 Best Local Similarity: 23.13% Mismatches: 151  
 Query Match: 6.35% Indels: 110  
 DB: 4 Gaps: 21  
 US-09-889-746-2 (1-400) x US-08-778-570B-5 (1-3660)  
 Qy 10 LeuSerLeuProLeuLeuSerValAla-----ValThrGlnGlnLeu 23  
 Db 322 TTGTCTTGGTCTGCTTAAACATCAAGCGAGTGGCACTGGCAACACACGCGCGATAAG 381  
 Qy 24 TyrAlaGlnProAsnGlnSerLeuProThrValGlnLeuGluProValIleThrIle 43  
 Db 382 GCGAGGCAACGATTAAGCAAACTTGTGTTGCTTGGATGAACCTGTGTAACAGCG 441  
 Qy 44 AspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrThrLysVal 63  
 Db 442 AAGAAAAAGCCCGTAA---GCCAAGAAAGTTACAGGCTTGTAAGTGTGCAAAACT 498  
 Qy 64 ILeTyrGlnGlnIleGlnGlnGlnAlaThrGlySerArgGlnLeuAlaAspValMet 83  
 Db 499 GCC---GAGACCATCAATTAAGAAACAGTCTAAACATTGAGACTTA----- 543  
 Qy 84 AlaGlnLeuIleProSerLeuGly---ValSerSerGlyThrThrSerAsnPheGlyGln 102  
 Db 544 ACAAGCTATGACCTTGCGCTGCTGTGTTGAGCAAGGTGCGGCGCAAGCTCAGGCTAT 603  
 Qy 103 ThrMetHisGly-----ArgGlnValGlnPheLeuLeuAsnGlyVal----- 116  
 Db 604 TCTATTGCGTATGATGATTAATAATCGTGTGGCGGTATGCTGATGATGATCAATCAAGCC 663  
 Qy 117 -----ProLeuThrGlySerArg---AspIleSerArgGlnLeu 128  
 Db 664 CAGCACTATGCCCTTACAGAGCCCTGTGGGAGCAAAATTATGCGGAGGTGGGCAATC 723

QY 129 AnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyAla---ThrSer 147  
DB 724 AACGAATAGAAATACGAAATGTCGCTCGGTAGATTAGTAAAGGTGCAAAATCAAGT 783  
QY 148 IleTyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLysSer----- 164  
DB 784 GAATACGGCTCTGGGCATATATCTGGCTCTGGCAATTTGTTACCAAAACCCCGCATGAC 843  
QY 165 -----AspLeuGluGluGlnPheGluThrArg----- 174  
DB 844 ATCATCAAGATGTTAAAGATTGGGCGGTGAGACCAACCAACCGCTATGCCAGTAAAT 903  
QY 175 -----IleGlyValHisGlySerLysLeuSerGluGlyIle--- 187  
DB 904 AACGCATGGTTAATTTCTGTGCAGCAGCAGGCAAGGCAGGTCTTTTAGCGGTCTATC 963  
QY 188 -----GlyTyrGlnValGly 192  
DB 964 ATCTACACGCGCGGTGTCAGAAATACAAGGCACATGATGATGCCTTACGGGTAGC 1023  
QY 193 GlnSer-----ValAla 196  
DB 1024 CAAAGTTTATAGAGCGGTGCAACCACTGACCAAAATACCAAAATTTTAAATAGCA 1083  
QY 197 GlyValSerGluAsnGlyAsnVal-----LeuAla 206  
DB 1084 AATGAATGTCCAATGGTAAATATATGAGCGGTGTGCTGGCGGTCAAAACCAACTCCAA 1143  
QY 207 ArgLeuAspValAspTyrArgThrThrGlyGlyAlaPheAspAlaAsnGly---LysArg 225  
DB 1144 GCTAGCCCAACCAATGTGCGTGAATAGTCAATGTCAAAAGATTATACAGGTCTTACCGC 1203  
QY 226 IleAlaProGluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsnThr 245  
DB 1204 CTTATCCCAACCCACTC-----ACCAAGACAGCAATCTTACTGCTCGCCCA 1254  
QY 246 AsnValAspTyrGlnLeuAspLysGlnAsnIleAsn-----LeuAlaLeuThrHis 263  
DB 1255 GGT-----TATCAGCTAAACGATATAGCACTATGTCGGGTGTGTATGAAATACCAAA 1308  
QY 264 TyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeu 283  
DB 1309 CAAACTAGCCCATGCAAGATAAACCCTGCTGCTTAT-----CTGAGGTTTCAAT 1359  
QY 284 PheGlyGluLysProSerLeuAsn-----AlaIleLysGlyLeu 296  
DB 1360 GACATTGAAAAATCAAGGCTCAGCAACCATGCGCAAGCCAAATGGCTATTATCAAGGCAAT 1419  
QY 297 SerLeuSerGluGlnProLysThr-----ThrLysSerThrPheAsnIleAsnTyr 313  
DB 1420 AACCTTGTGTAACGATTCGTGATGCCATTGGCGCAATTCAGGTTATGGCATCAACTAT 1479  
QY 314 HisHis 315  
DB 1480 GCTCAT 1485

RESULT 33  
US-09-584-5  
Sequence 5, Application US/09059584  
Patent No. 6440701

GENERAL INFORMATION:  
APPLICANT: Myers, Lisa E  
APPLICANT: Schryvers, Anthony B  
APPLICANT: Harkness, Robin E  
APPLICANT: Loonsmore, Sheena M.  
APPLICANT: Du, Run-Pan  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Klein, Michel H  
TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
NUMBER OF SEQUENCES: 60  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney

STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/059,584  
FILING DATE: 14-APR-1998

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/778,570  
FILING DATE: 03-JAN-1997

CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24973  
REFERENCE/DOCKET NUMBER: 1038-794  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163

INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3660 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-059-584-5

Alignment Scores:  
Pred. No.: 0.000115 Length: 3660  
Score: 130.00 Matches: 93  
Percent Similarity: 35.07% Conservative: 48  
Best Local Similarity: 23.13% Mismatches: 151  
Query Match: 6.35% Indels: 110  
DB: 4 Gaps: 21

US-09-889-746-2 (1-400) x US-09-059-584-5 (1-3660)

QY 10 LeuSerLeuProLeuLeuSerValAla-----ValThrGlnGlnLeu 23  
DB 322 TTGCTTTGGTCTGCTTAACATCAGCAGGTGGCAACCAACCAACCGCCGATAG 381  
QY 24 TyrAlaGlnProAsnGluSerLeuProThrValGluLeuGluProValIleThrIle 43  
DB 382 CGGAGGCAACAGATAAGACAAACCTTGTGTTGTTGGATGAAACTGTTGTAACGCG 441  
QY 44 AspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrThrLysVal 63  
DB 442 AAGAAAACCGCCGTAAA---GCCAACGAAGTTACAGGGCTTGGTAAGGTGTCAAACT 498  
QY 64 IleTyrGluGluGlnIleGlnGluGlnAlaThrGlySerArgGlnLeuAlaAspValMet 83  
DB 499 GCC---GAGACCATCAATAAAGAACAGTGCTAAACATTCGAGACTTA----- 543  
QY 84 AlaGlnLeuIleProSerLeuGly---ValSerSerGlyThrThrSerAsnPheGlyGln 102  
DB 544 ACACGCTATGACCTGGCATTGCTGTGTGTGAGCAAGTCTGTGGGGGCAAGCTCAGGCTAT 603  
QY 103 ThrMetHisGly-----ArgGlnValGlnPheLeuLeuAsnGlyVal----- 116  
DB 604 TCTATTCTGTGTATGATAAAAAATCGTGTGGCGGTATTGTTGTTGATGGCATCAATCAAGCC 563  
QY 117 -----ProLeuThrGlySerArg---AspIleSerArgGlnLeu 128  
DB 664 CAGCACTATGCCCTACAAAGGCCCTGTGGCAGCAAAAAATATTGCCGAGGTGGGCAATC 723  
QY 129 AsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyAla---ThrSer 147

Db 724 AACGAATGATACGAAATGCGCTCGCTTGAATTAAGTCAAAATTCAGT 783  
 Qy 148 ILeTyrGlySerGlyAlaThrGlyLeuLeaAnIleValThrIysSer----- 164  
 Db 784 GAATACGGCTCTGGGSCATTAATCTGGCTCTGGCATTTGTTACAAACCGCCATGAC 843  
 Qy 165 -----AspleuGluGluGluGluPheGluThrArg----- 174  
 Db 844 ATCATCAAGATGTTAAAGATTGGGCGGTCAGACCAAAACCCCTATGCACTAAAAAT 903  
 Qy 175 -----IleGlyValHisGlySerIlySerSerGluGlyIle----- 187  
 Db 904 AACGATGGGTTAATCTGTGGCAGACGAGCAAGGCAAGGTTCTTTAGCGGCTTATC 963  
 Qy 188 -----GlyTyrGluValGly----- 192  
 Db 964 ATCTACACGACCGCGGTGTGTAAGATACAGGACATGATGATGCTATCAGGGTAGC 1023  
 Qy 193 GlnSer-----ValAla 196  
 Db 1024 CAAAGTTTGTATAGACGGGTGCGACCACTGACCAAAATTAACCAAAATTTTAAATAGCA 1083  
 Qy 197 GlyValSerGluAsnGlyVal-----LeuAla 206  
 Db 1084 AATGAATGCGCATGCTTAATTAAGAGCGGTGCTGCTGGCGTCAACCAAACTCCAA 1143  
 Qy 207 ArgLeuAspValAspTyrArgThrThrGlyAlaPheAspAlaAsnGly--LysArg 225  
 Db 1144 GCTAAGCCCAACCAATGCTGTATAGCTCAATGTCAAAGATTATACAGTCTTAACCGC 1203  
 Qy 226 ILeAlaProGluProAlaGlnThrAspLysGlnAspSerIysSerLeuValAsnThr 245  
 Db 1204 CTATATCCCAACCCATC-----ACCCAGACAGCAAACTTACTGCTTCCGCCA 1254  
 Qy 246 AsnValAspTyrGluLeuAspAspLysGlnAsnIleAsn-----LeuAlaLeuThrHis 263  
 Db 1255 GGT-----TATCAGCTAAACGATAGCACTATGCGGTGTGTATGATGAACACCAAA 1308  
 Qy 264 TyrAsnAspLysGlnAspThrAspTyrAlaProAspTyrGlyValAsnArgLeuAlaValLeu 283  
 Db 1309 CAATATGACCCATGCAAGATTAACCCGTCCTCTTAT-----CTGACGGTTCAT 1359  
 Qy 284 PheGlyGluLysProSerLeuAsn-----AlaIleGlyLeu 296  
 Db 1360 GACATGAAATAATCAAGGCTCAGCAACATGCGCAAGCAATGCTATATCAAGGCAT 1419  
 Qy 297 SerLeuSerGluGlnProLysThr-----ThrLysSerThrPheAsnIleAsnTyr 313  
 Db 1420 AACCTTGTAAGCATGCTGTATGTCATGCGCAAAATTCAGTTATGCGATCAACTAT 1479  
 Qy 314 HisHis 315  
 Db 1480 GCTCAT 1485  
 RESULT 34  
 US-09-453-702B-57/c  
 ; Sequence 57, Application US/09453702B  
 ; Patent No. 6365723  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Blattner, Frederick R.  
 ; Burland, Valerie  
 ; Perna, Nicole T.  
 ; Plunkett, Guy  
 ; Welch, Rod  
 ; TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157  
 ; NUMBER OF SEQUENCES: 265  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Quarles & Brady  
 ; STREET: 1 South Pinckney Street  
 ; CITY: Madison  
 ; STATE: WI  
 ; COUNTRY: US  
 ; ZIP: 53701-2113

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Word Perfect 8.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/453,702B  
 ; FILING DATE: 03-Dec-1999  
 ; CLASSIFICATION: <unknown>  
 ; PRIORITY APPLICATION DATA:  
 ; APPLICATION NUMBER: 60/110,955  
 ; FILING DATE: 04-Dec-1998  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Seay, Nicholas J.  
 ; REGISTRATION NUMBER: 27386  
 ; REFERENCE/DOCKET NUMBER: 960296,95017  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (608) 251-5000  
 ; TELEFAX: (608) 251-9166  
 ; INFORMATION FOR SEQ ID NO: 57:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 87563  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; MOLECULE TYPE: DNA (genomic)  
 ; TOPOLOGY: linear  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 57:  
 US-09-453-702B-57  
 Alignment Scores:  
 Pred. No.: 0.0209 Length: 87563  
 Score: 130.00 Matches: 58  
 Percent Similarity: 38.93% Conservative: 37  
 Best Local Similarity: 23.77% Mismatches: 79  
 Query Match: 6.35% Indels: 70  
 DB: 4 Gaps: 11  
 US-09-889-746-2 (1-400) x US-09-453-702B-57 (1-87563)  
 Qy 40 ValIleThrIleAspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHis 59  
 Db 49913 GTGATGATTGTCTCGGCATCCGCG-----TATGAGAAAACCTGACTAACCCAGCCGCC 49860  
 Qy 60 ThrThrLysValIleTyrGluGluGlnIleGlnGluAlaThrGlySerArgGlnLeu 79  
 Db 49859 AGTGTTCTGTGATTAAGCCAGAGGATTCAG-----TCCAGCCATGAC 49815  
 Qy 80 AlaAspValMetAlaGlnLeu-----IleProSerLeuGlyValSerSerGlyThrThr 97  
 Db 49814 CAGATCTGCGGAGGCTCTGATGATCAGTAGAGGGGTGGATGTTGAAAGTGTACGGGT 49755  
 Qy 98 SerAsnPheGly-----GlnThrMetHisGlyArgGlnValGlnPheLeu 112  
 Db 49754 AAAACCGAGGGCTGGAATTCAGATCCGAGGAAATGCCAGCATTAACCGCTGATATCG 49695  
 Qy 113 LeuAsnGlyValProLeuThrGlySerArgAspIleSer---ArgGlnLeuAsnSerIle 131  
 Db 49694 ATTGATGTTGTTCTGACGAGGCGGAGACAGTACCTCCCAACGGTTTCTTCCCATG 49635  
 Qy 146 ThrSer---IleTyrGlySerGlyAlaThrGlyLeuLeaAnIleValThrIysSer 164  
 Db 49574 ATGTCCACACTGTATGCTCTATGATGATGAGGCGGTGTGTATATCATTCACAGAAAG 49515  
 Qy 165 AspleuGluGluGluGluPheGluThrArgIleGlyValHis----- 178  
 Db 49514 AATGACAGCAAAATGCTCTTCCGTCATGACAGGCTTAATCTGACAGAAACCAAA 49455  
 Qy 179 -----GlySerIlySerSerGluGlyIle 187

Db 49454 TCGGGTAACAGCAGCCAGTTTAAATTTCTGGAGCAGTGGTCCCTTGTGGATGATTCGTG 49395  
Qy 188 GlyTyrGlnVal-----GlyGlnSerValAlaGlyVal--- 198  
Db 49394 AGCTGTCAGGTACGGCGTAGCACACACAGCGTTCAGGGTTTCATCGGTACATCATCTGAGC 49335  
Qy 199 -----SerGlnAsnGlyAsnValLeuAla 206  
Db 49334 GATACAGCAGCCAGCGGTATTCCTTATCCACGGAGTCACAGAAATTATAATCTTGGTGCA 49275  
Qy 207 ArgLeuAsp-----ValAspTyrArgThrThrGly 216  
Db 49274 CGTCTGTGACTGGAAGCGTCGGAGCAGGATGCTCTGCTGTTGATGATGATACCCCGG 49215  
Qy 217 GlyAlaPheAsp 220  
Db 49214 CAGCGTTATGAT 49203

RESULT 35  
US-09-307-973A-1  
; Sequence 1, Application US/09307973A  
; Patent No. 6432686  
; GENERAL INFORMATION:  
; APPLICANT: BULTIUS, BEN  
; APPLICANT: GATENBY, ANTHONY A.  
; APPLICANT: TRIMBUR, DONALD E.  
; APPLICANT: WHITED, GREGORY  
; TITLE OF INVENTION: METHOD FOR THE PRODUCTION OF  
; 1,3-PROPANEDIOL BY RECOMBINANT  
; TITLE OF INVENTION: ORGANISMS COMPRISING GENES FOR  
; TITLE OF INVENTION: VITAMIN B12 TRANSPORT  
; NUMBER OF SEQUENCES: 25  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: E. I. DUPONT DE NEMOURS AND COMPANY  
; STREET: 1007 MARKET STREET  
; CITY: WILMINGTON  
; STATE: DELAWARE  
; COUNTRY: U.S.A.  
; ZIP: 19898  
; ADDRESSEE: GENENCOR INTERNATIONAL, INC.  
; STREET: 925 PAGE MILL ROAD  
; CITY: PALO ALTO  
; STATE: CALIFORNIA  
; COUNTRY: U.S.A.  
; ZIP: 94304-1013  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: DISKETTE, 3.50 INCH  
; COMPUTER: IBM PC COMPATIBLE  
; OPERATING SYSTEM: MICROSOFT WINDOWS 95  
; SOFTWARE: MICROSOFT OFFICE 97  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/307,973A  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/085,190  
; FILING DATE: JUNE 30, 1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: FLOYD, LINDA AXAMETHY  
; REGISTRATION NUMBER: 33,692  
; REFERENCE/DOCKET NUMBER: CL-1245  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 302-892-8112  
; TELEFAX: 302-773-0164  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1845 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO

ANTI-SENSE: NO  
US-09-307-973A-1  
Alignment Scores:  
Pred. No.: 6.46-05 Length: 1845  
Score: 128.00 Matches: 75  
Percent Similarity: 36.64% Conservative: 32  
Best Local Similarity: 25.68% Mismatches: 99  
Query Match: 6.25% Indels: 86  
Gaps: 15

US-09-889-746-2 (1-400) x US-09-307-973A-1 (1-1845)

Qy 51 AlaAsnArgGlnThrGlnMetProHisThr-----ThrLysValIleTyrGlu 66  
Db 94 GCTAACCGTTTTGAACAGCCGCGCAGCATGTGCTTGACCAACACCGTGTGTGACCGGT 153  
Qy 67 GluGlnIleGlnGluAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeu 86  
Db 154 CAGGATATCGACCGCTGCGCAG-----TCGACCTCGGTCAATGATGTCTGCGCGTCTT 207  
Qy 87 -----IleProSerLeuGlyValSerSerGlyThrThrSerAsnPheGlyGln 102  
Db 208 CCGGCGCTCGATATCACCCAAACGCGGTTTCAGGTACGTCTCATCTATTTTATTTCGC 267  
Qy 103 ThrMetHisGlyArgGlnValGlnPheLeuLeuAsnGlyValProLeu----- 118  
Db 268 GGTACAAATGCCAGTCACTGTGTGTTGTTAATGATGCGTACGCTGAATCTCGCGGG 327  
Qy 119 ---ThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAsnProAsnGlnValAla 137  
Db 328 GTGAGTGTGTTCTGCCGACCTTAGCCAGTTCCTTATGTCGCTT-----GTCGAG 375  
Qy 138 ArgIleGluValLeuSerGlyAlaThrSer---IleTyrGlySerGlyAlaThrGly 156  
Db 376 CGTGTGTAATATATCCGTGGCGCGCTCCGCTGTTTATGTTCCGATGCAATAGCGGG 435  
Qy 157 LeuIleAsnIleValThrLysSerAspLeuGluGluGlnPheGluThrArgIleGly 176  
Db 436 GTGTTGAATATATCATCAGCAGCGCGAT-----GAACCCGGAACGGAA 477  
Qy 177 ValHisGlySerLysLeuSerSerGluGlyIleGlyTyrGlnValGly---GlnSerVal 195  
Db 478 ATTTCCAGGAGGTGGGGAAGCAATAGTTATCAGAACTATGATGTTCTTACGCGCAACAA 537  
Qy 196 AlaGlyValSerGluAsnGlyAsnValLeuAlaArgLeuAspValAspTyrArgThrThr 215  
Db 538 CTGGGGGATACAGACACGCGGTAACTGTTGGC-----GATTATGCCCATACT 585  
Qy 216 GlyGly-----AlaPheAspAlaAsnGlyLysArgIleAlaProGluProAla 231  
Db 586 CATGTTATGATGTTGTTGCTTAAACCGCAACGCA-----GCG 630  
Qy 232 GlnThrAspLysGlnAsp-----SerLysSerLeu----- 241  
Db 631 CAGACAGATAACGATGTTTAAAGTAAACGCTTATGCGCGCTGGAGCATACTTT 690  
Qy 242 -----SerValAsnThrAsnValAsp 248  
Db 691 ACTGATGCTCGGAGCGGCTTTGTGCGCGCTATGGCTATGATAACCGTACCAATTATGAC 750  
Qy 249 -----Trp 249  
Db 751 GCGTATTATTTCTCCCGGTTTCCAGCTTGTGCTGATACCCGTAACCTCTATAGCCAAAGTTGG 810  
Qy 250 GlnLeuAspAspLysGlnAsnIleAsnLeu-----AlaLeuThrHisTyrAsn 265  
Db 811 GACGGCGGCTCGGCTATAACCGGCAACTGATTAATACAACTCATTACCAGCTATAGC 870  
Qy 266 AspLysGlnAspThrAspTyrAlaProAspTyrGly 277  
Db 871 CATAGCAAGATTACAACTACCGATCCCGATTATGTT 906



```

RESULT 36
US-08-537-361E-7
; Sequence 7, Application US/08537361E
; Patent No. 6121037
; GENERAL INFORMATION:
; APPLICANT: Stojiljkovic, Igor
; APPLICANT: So, Magdalene
; APPLICANT: Hwa, Vivan
; APPLICANT: Heffron, Fred
; APPLICANT: Nassif, Xavier
; TITLE OF INVENTION: No. 6121037el Bacterial Hemoglobin Receptor
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESS: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/537,361E
; FILING DATE: 02-OCT-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6121037nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 94,784-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2378 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2370
; US-08-537-361E-7

Alignment Scores:
Pred. No.: 9,71e-05 Length: 2378
Score: 128.00 Matches: 95
Percent Similarity: 34.24% Conservative: 44
Best Local Similarity: 23.40% Mismatches: 151
Query Match: 6.25% Indels: 116
DB: 3 Gaps: 22

US-09-889-746-2 (1-400) x US-08-537-361E-7 (1-2378)
QY 12 LeuProLeuLeuSerValAlaValThrGlnGlnLeuTyrAlaGlnPro----- 27
DB 19 CTTCTCTATT-----GCCGGCTGCTGCGCAGTATTTTCGGCAATCGTCTTGAGAGCG 72
QY 28 -----AsnGluSerLeuProThr-----ValGluLeuGlnProValIleThr 42
DB 73 GATGAAGCTGCAACCAACACACACCGCTTAAAGCAGAGATTAAGAAGTCCGCTTAA 132
QY 43 IleAspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrThrLys 62
DB 133 GACCACTTAATGCGCTGCAACCGTGGAACGTGTC-----AACCTCGGC 177
QY 63 ValIleTyrGlnGlnGlnIleGlnGlnGlnIleThrGlySerArgGlnLeuAlaAspVal 82
DB 178 CGCATTCACAGAAATGATACGCGACACAAAGACTTGTGTGCTTACTCCACCGAGCTC 237

```

```

QY 83 MetAlaGlnLeuIleProSerLeuGlyValSerSerGlyThrThrSerAsnPheGly--- 101
DB 238 -----GCTTGAAGCGATTACCGCGCGCATCAAAAAGGCTTT 273
QY 102 -----GlnThrMetHisGlyArgGlnValGlnPheLeuLeuAsnGlyValProLeuThr 119
DB 274 GCTGTGCGCGCGGTGAAAGCAACCGTGTGCTGTGCACATTGACGGGTGAGCTGTGCT 333
QY 120 GlySerArgAspLysSer-----ArgGlnLeuAsn-----Ser 130
DB 334 GATTGGAAGAAACTCACTGTATGCACTTATGGCACTTCAACAGCTCGCGCTGTCT 393
QY 131 IleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyAlaThrSerIle---Tyr 149
DB 394 ATGACACCCGAACCTGTGCGCAACATCGAATTCGGAGAGGGCGCTGACTTTCAATACC 453
QY 150 GlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLysSerAsp----- 165
DB 454 GGTACCGCGGCATTGCTGTGCGCGCTGAATTACAAACCTCGCAAGACATGATTGTCTG 513
QY 166 LeuGlnGlnGlnGlnPhe-----GluThrArg----- 174
DB 514 TTGACAGACAGCAATTGCGCGTGAATGATGAAAAACGTTACAGACGCGCAACCGCGAA 573
QY 175 -----IleGlyValHisGlySerLysLeuSerSerGlyIle--- 187
DB 574 TGGACAAATACACTGCGTTTGTGCTGTGAGCAACGCGGTGATGCGCGCTTGTGCTGAT 633
QY 188 GlyTyrGlnValGlyGlnSerValAlaGlyValSerGluAsnGlyAsnValLeuAlaArg 207
DB 634 TCGCAACGTGCGCGCTCATGACACCAAGCGCGCGCGCTGCG- 678
QY 208 LeuAspValAspTyrArgThrThrGly-----GlyAlaPheAspAlaAsnGlyLys 224
DB 679 -----TATCCGCTAGAGGTGCTGCGACCGGACGACATTAATCCGTGCTGCTCA 726
QY 225 ArgIleAlaProGlnProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsn 244
DB 727 CGCGGTATCCCGATCCGTCGCAAA-----CAAAATACCAACACTTCTTG 771
QY 245 ThrAsnValAspTyrGlnLeuAspAspLysGlnAsnIleAsnLeuAlaLeuThrHisTyr 264
DB 772 GGTAAAGATTGCTTATCAATCAACACACACACCGCATCGC-----CCATCGTT 822
QY 265 AsnAspLysGlnAspThrAspTyr----- 272
DB 823 AACGCGCAGCAGGCGCATTAATTAACAGATTGAAGCTTATTAACCTGACCGCTTCTCC 882
QY 273 -----AlaProAspTyrGlyAsnArgLeu---AlaValIleuPheGlyGlu---Lys 287
DB 883 TGGCGCGAAAGCGGATGACCTTAACACAGACGCGCAATGCCAATCTTTTACAAAGAGAC 942
QY 288 ProSerLeuAsnAlaIleLysGlyLeuSerLeuSerGlnGlnProLysThrThrLys--- 306
DB 943 CCGATTCAAATTGCGTGTGCTTTGAAGCGGATTCGATTATACAGACAAACAAGTG 1002
QY 307 -----SerThrPheAsnIleAsnTyr 313
DB 1003 GCGCGGCTTAACAAAGAGCTGTCCGACGAGATTAATTCACCTTGACGCGCAACTAT 1062
QY 314 HisIleAspAspLeuTyrGlyAsnThrIleAsnThrAsnAlaTyrTyrArgArgGluLys 333
DB 1063 AATCAGAGGATTTG-----GAGATATATACACCGCAGCATGAGC 1104
QY 334 GlyArgPheTyrProPhe 339
DB 1105 ACCCGATTCAAAAGTTT 1122

RESULT 37
US-08-817-707-7
; Sequence 7, Application US/08817707
; Patent No. 6277382

```

```

; GENERAL INFORMATION:
; APPLICANT: Stojiljkovic, Igor
; APPLICANT: So, Magdalene
; APPLICANT: Hwa, Vivian
; APPLICANT: Heffron, Fred
; APPLICANT: Nassif, Xavier
; TITLE OF INVENTION: No. 6277382el Bacterial Hemoglobin Receptor
; TITLE OF INVENTION: Genes and Uses
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESS: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 South Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,707
; FILING DATE: 19-AUG-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6277382nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 94,784-J
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2378 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2370
US-08-817-707-7

```

```

Alignment Scores:
Pred. No.: 9,71e-05 Length: 2378
Score: 128.00 Matches: 95
Percent Similarity: 34.24% Conservative: 44
Best Local Similarity: 23.40% Mismatches: 151
Query Match: 6.25% Indels: 116
DB: 4 Gaps: 21

US-09-889-746-2 (1-400) x US-08-817-707-7 (1-2378)

QY 12 LeuProLeuSerValAlaValThrGlnGlnLeuTyrAlaGlnPro----- 27
Db 19 CTTCTATT-----GCCGCGTTCGCGCAGTATTTCGGCAATCCGCTTGGCAGCG 72
QY 28 -----AsnGluSerLeuProThr---ValGluLeuProValValIleThr 42
Db 73 GATGAAGTGCACCGAACCACCCGTTAAGCAGAGATAAAGAAAGTGGCGGTAAA 132
QY 43 IleAspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrThrLys 62
Db 133 GACCAGCTTAATCGCGCTGCAACCGTGGACCGTGC-----AACCTCGGC 177
QY 63 ValIleTyrGluGluGlnIleGlnGluGlnAlaThrGlySerArgGlnLeuAlaAspVal 82
Db 178 CGCATTCACAGGAATGATACGCGACAAACAAAGACTTGGTGGGTACTCCACCGACGTC 237
QY 83 MetAlaGlnLeuProSerLeuGlyValSerSerGlyThrThrSerAsnPheGly--- 101
Db 238 -----GGCTTGAGCGATAGCGCGGCATCAAAAGAGCTTT 273

```

```

QY 102 -----GlnThrMetHisGlyArgGlnValGlnPheLeuLeuAenGlyValProLeuThr 119
Db 274 GCTGTGCGCGCGTGGAAAGCAACCGTGTGTCAGCATTTAGCGGTGAGCGCTGCT 333
QY 120 GlySerArgAspIleSer-----ArgGlnLeuAen-----Ser 130
Db 334 GATTCCGAAGAAACTCACTGTATGTCAGCTTATGGCAACTTCAACAGCTCGCGCTGCT 393
QY 131 IleAenProAenGlnValAlaAargIleGluValLeuSerGlyAlaThrSerIle---Tyr 149
Db 394 ATCGACCCCGAACTCGTGCACACATCGAATCGGAAGGCGCTGACTCTTTCATACC 453
QY 150 GlySerGlyAlaThrGlyGlyLeuIleAenIleValThrLysSerAsp----- 165
Db 454 GGTAGCGCGCATTTGGTGGCGCGTGAATTAACCAACCCCTCAAGGACATGATTGCTG 513
QY 166 LeuGluGluGlnPheGluThrArgIleGlyValHisGlySerLysLeuSerSerGlu 185
Db 514 TTGGACGACAGGCAATTCGGCGTGATGATGAAACAGGTTACAGCAGCGCAACCGCAA 573
QY 186 -----GlyIleGlyTyrGlnVal----- 191
Db 574 TGGACAAATACACTCGGTTTCGGTGTGAGCAACGACCGGTGGATCGCGCTTGTGTAT 633
QY 192 -----GlyGlnSerValAlaGlyValSerGluAenGlyAenValLeuAlaArg 207
Db 634 TCGCAACGTCGCGTCATGAGACGAAAGCGGCGGAGCGTGGC----- 678
QY 208 LeuAspValAspTyrArgThrGly-----GlyAlaPheAspAlaAenGlyLys 224
Db 679 -----TATCCGCTAGAGGCTCTGGCAGCGAGCAATTTATCCGTGGTTCGTCA 726
QY 225 ArgIleAlaProGluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAen 244
Db 727 CGCGGTATCCCTGATCCCTCCAAA-----CACAAATACCAACACTTCTTG 771
QY 245 ThrAenValAspTyrGlnLeuAspAspLysGlnAenIleAenLeuAlaLeuThrHisTyr 264
Db 772 GGTAAAGATTGCTTATCAATCAACGACGACGACCGCATCGGC-----CCATCGTTT 822
QY 265 AsnAspLysGlnAspThrAspTyr----- 272
Db 823 AACGCCAGCAGCGGCATTAATTACACGATTAAGAGTCTTATAACCTGACCGCTTCTTCC 882
QY 273 -----AlaProAspTyrGlyAsnArgLeu---AlaValLeuPheGlyGlu---Lys 287
Db 883 TGGCGCGAAGCGCGATGAGTAAACAGACGCGCGCATGCGCAACCTCTTTTACGATGGAGC 942
QY 288 ProSerLeuAenAlaIleLysGlyLeuSerLeuSerGluGlnProLysThrThrLys--- 306
Db 943 CCTGATTCAAAATTGGCTGCTGCTTTGAAGCGGATTTTCGATTATCAGACCAACAAAGTG 1002
QY 307 ----- 313
Db 1003 GCGGCGGTAAACAAACAAAGCTCGTTCGCCAGCGGATTTATCCCTGACGCGCAACTAT 1062
QY 314 HisHisAspAspLeuTyrGlyAenThrIleAenThrAenAlaTyrTyrArgGluLys 333
Db 1063 AATCAGAGGATTG-----GAGAATATATACACCGGACGATGGAC 1104
QY 334 GlyArgPheTyrProPhe 339
Db 1105 ACCCGATTCAAAACGTTTT 1122

```

```

RESULT 38
US-09-221-017B-462/c
; Sequence 462, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120

```

```

CORRESPONDENCE ADDRESS:
ADDRESS: MORRISON & FOERSTER
STREET: 755 PAGE MILL ROAD
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/221.017B
FILING DATE: 23-DEC-1998
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PP1182
FILING DATE: 31-DEC-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP1546
FILING DATE: 30-JAN-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PP2911
FILING DATE: 09-APR-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/AU98/01023
FILING DATE: 10-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Montoy, Gladys H
REGISTRATION NUMBER: 32,430
REFERENCE/DOCKET NUMBER: 27340-20021.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-813-5600
TELEX: 650-494-0792
TELEX: 706141
INFORMATION FOR SEQ ID NO: 462:
SEQUENCE CHARACTERISTICS:
LENGTH: 2760 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: UNKNOWN
ORIGINAL SOURCE:
ORGANISM: FORIPHROMONAS GINGIVALIS
FEATURE:
NAME/KEY: misc feature
LOCATION: 1...2760
US-09-221-017B-462
Alignment Scores:
Pred. No.: 0.000162 Length: 2760
Score: 127.00 Matches: 113
Percent Similarity: 35.22% Conservative: 61
Best Local Similarity: 23.87% Mismatches: 165
Query Match: 6.20% Indels: 157
Dels: 4 Gaps: 25
US-09-889-746-2 (1-400) x US-09-221-017B-462 (1-2760)
QY 34 ValGluLeuGluProValValIleThrIleAspLysSerGlyMetAlaLeuAlaAspArg 53
Db 1744 ATCAATCTGACAGATCGGATTCG-----GCCAACCGC 1709
QY 54 -----IleThrGlnMetProHisThrThrLysValIleLeuGluGlnIleGln 70
Db 1708 GAAGTACGCGTTCGCGCTCTGCTCTACTCTGTAATGATATGAAAGAAAGCTTC 1649
QY 71 GluGlnAlaThrLysSerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeu 90
Db 1648 TCGCAAGTCAATGCTTCAACTGCTGCTCAAGGCTTGCATTCAG-----CCGGAGATT 1595

```

```

QY 91 GlyAlaSerSerGlyThrThrSer---AspPheGlyGlnThr-----MetHis 105
Db 1594 CGGTAGAGAAACACTGTCAGAACTGTGTTTCATCAAGTTCGATCAATGAGCTGAT 1535
QY 106 GlyArgGlnValGlnPheLeuLeuGlnGlyValProLeuThrGlySerArgAspIleSer 125
Db 1534 GGTTCGTTATGACAGATCCCTCATCGACAGCCGTCCTCATGAGTGCCTCCGCTGT 1475
QY 126 ArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGlnValLeu---SerGly 144
Db 1474 TACGCTCTGAGACAGATCCCTGCAATATGATCGAACGGTGTAGAGTATGAGTGTGA 1415
QY 145 AlaThrSerIleLeuGlySerGlyValThrGlyGlyLeuIleAsnIleValThrLysSer 164
Db 1414 GGATCGCGCTTACGCTTCTCTCTCTATTCGCGAGTGTAAATATCATCAAGAA 1355
QY 165 AspLeuGluGlu-----GluGlnPhe 171
Db 1354 CCTCTCACAATTCCTTTCACATTCATGATGATCTGAGCTTACCGGTTTCAGCAAGCTG 1295
QY 172 GluThrArgIleGlyValHisGlySerLysLeuSerSerGluGlyIleGlyThrGlnVal 191
Db 1294 GATTAACAACACCACTTCATGCTTCATCGACGATGACAC-----CGTCC 1244
QY 192 GlyGlnSerValAlaGlyValSerGluAsnGlnAsn----- 203
Db 1243 GGTGCGATGTATTTGGCGGAGCTGCTTACCGCAACCATTTGGAGATGCTAACATGACGT 1184
QY 204 -----ValLeuAlaArgLeuAspVal----- 210
Db 1183 TATTCGGAATGGTAAATAGATGCCGCTGAGAGCGCATTTCTTATTTGCGCTTG 1124
QY 211 ---AspTyrArgThrThrGlyGlyAlaPheAspAla-----AsnGly 223
Db 1123 AGCGACTACAGCAAAATTGACGGAGAGTTTCACAGATCAGTAATTCGCGGTGGC 1064
QY 224 LysArgIleAlaProGluPro-----AlaGlnThrAsp----- 234
Db 1063 GATCGTATGATTTGGCTCTCTACGATGAGGTGAGTACGTGAACCAATGACATAGCGTA 1004
QY 235 -----LysGlnAsp-SerLysSerLeuSerValAsnThrAsnValAspTr 249
Db 1003 TTTAGCGAAACTGTAATATGATCTCTCTTCCAACTATAAACACACACTTCACGCT 944
QY 249 PqlnLeuAspAspLysGlnAsnIleAsnLeuAlaLeuThrHisTyrAsnAspLysGlnAs 269
Db 943 TATACTTCGGAAGATCGTAATGCAAGACCTATTACGGAGTATCGAGAGATT-GA 885
QY 269 PThrAspTyrAlaPro-----AspTyrGlyAs 278
Db 884 CGTCAATGGCCACCCCGGTGTACGGAAGGTAACCTATCCCTCAAGATCAATACGGCA 825
QY 278 PArgLeuAlaValLeuPheGlyGlu-----LysProSerLeuAsnAl 292
Db 824 TAATTATGCGCGACCAAGCAAGACATATATGGCGGTATCCAGTACAGTACGACTT 765
QY 292 allelysglyLeuSerLeuSerGluGlnProLysThrThrLysSerThrPheAsnIleAs 312
Db 764 GGACAAATTCCTCTCATCTTCGCA-----CTTTGTTCGAGCCGA 720
QY 312 nTyrHisHisAspAspLeu-----TTP----- 319
Db 719 ATATACGCTGATGAACTCAATGAGCTGATGCCATCTTTATGCGACAGCCGCGAGGA 660
QY 320 -----GlyAsnThr-----IleAsn-ThrAsn 327
Db 659 TCGCAATGGGAATACATTCCTCTATCCGGAATTTGATCAAGATATATCAACATCAG 600
QY 327 LAtyTyrArgArgGluLysGlyArgPheTyrPro-----PheValAlaP 342
Db 599 CCTATTGCGTCAAGCAATGGAATAAATGACAGATGAGACATCTTTGTTGGCTCGCT 540

```

QY 342 roPheSerIleAlaLysAlaLeuProIleLeuGlnSerMetAsnLeuProSerAlaThrL 362  
DB 539 GGACAGCATAGCAAGTCAAG-GATATGATTCTGAGT-----CCTGTACCACAC 490  
QY 362 eu-----AspAlaTyrThrLysAlaProGlnAlaAArgAlaTyr--- 374  
DB 489 TCGGTTTCAACGTGAATCCGGACATCAACTGCGCTACATATGCAAAAGGGTCCGCG 430  
QY 375 -----GlyValLeuGlnSerGluSerLysAlaG 384  
DB 429 CACCGCAGGTATTTCGATGAAGACTTCACAGTAGGGTGTAGCGGTGAG-----GCAC 376  
QY 384 luValLeuGlyArgValProAsnLeuAsnLysProLys 396  
DB 375 AGAAGATATTCAACGATCCGAACCTC---AAGCCTGAA 341

## RESULT 39

US-08-425-843-1  
; Sequence 1, Application US/08425843  
; Patent No. 6020154  
; GENERAL INFORMATION:  
; APPLICANT: Hansen, Eric J.  
; APPLICANT: Cope, Leslie D.  
; APPLICANT: Jarosik, Gregory P.  
; APPLICANT: Hanson, Mark S.  
; TITLE OF INVENTION: H. Influenzae HxuD and HxuC Genes, Proteins  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433  
; CITY: Houston  
; STATE: Texas  
; COUNTRY: USA  
; ZIP: 77210-4433  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/425,843  
; FILING DATE: Concurrently Herewith  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Parker, David L.  
; REGISTRATION NUMBER: 32,165  
; REFERENCE/DOCKET NUMBER: AMCY:012/PAR  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (512) 418-3000  
; TELEFAX: (713) 789-2679  
; TELEX: 79-0924  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4061 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA  
US-08-425-843-1

Alignment Scores:  
Pred. No.: 0.000778 Length: 4061  
Score: 123.50 Matches: 74  
Percent Similarity: 38.85% Conservative: 48  
Best Local Similarity: 23.57% Mismatches: 117  
Query Match: 6.03% Indels: 75  
DB: 3 Gaps: 15

US-09-889-746-2 (1-400) x US-08-425-843-1 (1-4061)

QY 7 PheGlnTrpLeuSerLeuProLeuSerValAlaValThrGlnLeuTyrAlaGln 26  
||||| ||||||| ||| :||| |||||

DB 118 TTTTCTAACTTCCTTCCATTGCAATTCGCAACACCTTAGTCAGACGAAATCGCGTACGCGAA 177  
QY 27 ProAsnGluSerLeuProThrValGlnLeuGluPro-----ValValIleThrIleAsp 44  
DB 178 -----TCGGTTGAATTAGACTCTATCAACGTTTATTTCGACACAGAGAT 219  
QY 45 LysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrThrLysValIle 64  
DB 220 CCAAGTAGGTTTGCCTTATACG----- 240  
QY 65 TyrGluGlnGlnIleGlnGlnAlaThrGlySerArgGlnLeuAlaAspValMetAla 84  
DB 241 ---CCAGAAAACAATCTAAGATAGTCTCTTTCCAGCAAGCCACTAGTGTTCAGCA 297  
QY 85 GlnLeu-----IleProSerLeuGlyValSerSerGlyThrThrSer----- 98  
DB 298 GCGTTAGAAGACATCTCCCAATGTTGATATTAGAGCGGTTCAAGAAGCATTCGCTCAAAA 357  
QY 99 ---AsnPheGlyGlnThrMetHisGlyArgGlnValGlnPheLeuLeuAsnGlyValPro 117  
DB 358 CCTAATATTCGAGGGTTAAGTGATTAATCGTGTGTGCAA---GTCAATTGATGGCGTG--- 411  
QY 118 LeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAsnPro---AsnGlnVal 136  
DB 412 ---AGACAAAATTTGATTTAGACATAGAGTCTTATTCTTCCATGTCACTCATC 468  
QY 137 AlaArgIleGluValLeuSerGly---AlaThrSerIleTyrGlySerGlyAlaThrGly 155  
DB 469 CAAGAAATTTGAAGTAATCAAGAGCAAGTAGTCTCTTATGGGTAGCGGTGATGGGT 528  
QY 156 GlyLeuIleAsnIleValThrLysSerAspLeuGluGlnPheGluThrArgIle 175  
DB 529 GGTGTGTGGCAATCGTACGCCAAATCTTTGGACTTATTGAAAAATAATGACAAATTC 588  
QY 176 GlyValHisGlySerLysLeuSerSerGluGlyIleGlyTyrGlnValGlyGlnSerVal 195  
DB 589 GGAGTT-----AAAATTCGCCAA-----GGTTATCAA----- 615  
QY 196 AlaGlyValSerGluAsnGlyAsnValLeuAlaArgLeuAspValAspTyrArgThrThr 215  
DB 616 -----ACTGCTAATAATTTATCGAAAGGAGATGTTCTGTATTTCGGCGCA 660  
QY 216 GlyGlyAlaPheAspAla-----AsnGlyLysArgIleAlaProGluProAlaGlnThr 233  
DB 661 AATGACAAATTCGATGTTCTTATTAGTGTCTTATAATAATGCGGATAATTTACGCACT 720  
QY 234 AspLysGlnAspSer-----LysSerLeuSerValAsnThrAsnVal 247  
DB 721 GGTAAAGGCAATAAGCTGAATATACCGCTATAACAGTTTGGGGGCTTAGCAAAATTC 780  
QY 248 AspTrpGlnLeuAspAspLysGlnAsnIleAsnLeuAlaLeuThrHisTyrAsnAspLys 267  
DB 781 GGTTCGCAAAATTAATGATGCTAAACCGCGTGAATTTATCCACCGCGAAACTCGTTTAAA 840  
QY 268 GlnAspThrAspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeuPheGlyGluLys 287  
DB 841 CAACACAGCA----- 849  
QY 288 ProSerLeuAsnAlaIleLysGlyLeuSerLeuSerGluGln 301  
DB 850 CCAAGCAATAATGAGTGGGAAACCAACTTACCAATGACAA 891

## RESULT 40

US-09-307-973A-2  
; Sequence 2, Application US/09307973A  
; Patent No. 6432686  
; GENERAL INFORMATION:  
; APPLICANT: BULTHUIS, BEN  
; APPLICANT: GATENBY, ANTHONY A.  
; APPLICANT: TRIMBUR, DONALD E.  
; APPLICANT: WHITED, GREGORY  
; TITLE OF INVENTION: METHOD FOR THE PRODUCTION OF  
; 1,3-PROPANEDIOL BY RECOMBINANT

TITLE OF INVENTION: ORGANISMS COMPRISING GENES FOR  
 TITLE OF INVENTION: VITAMIN B12 TRANSPORT  
 NUMBER OF SEQUENCES: 25  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: E. I. DUPONT DE NEMOURS AND COMPANY  
 STREET: 1007 MARKET STREET  
 CITY: WILMINGTON  
 STATE: DELAWARE  
 COUNTRY: U.S.A.  
 ZIP: 19898  
 ADDRESSEE: GENENCOR INTERNATIONAL, INC.  
 STREET: 925 PAGE MILL ROAD  
 CITY: PALO ALTO  
 STATE: CALIFORNIA  
 COUNTRY: U.S.A.  
 ZIP: 94304-1013  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: DISKETTE, 3.50 INCH  
 COMPUTER: IBM PC COMPATIBLE  
 OPERATING SYSTEM: MICROSOFT WINDOWS 95  
 SOFTWARE: MICROSOFT OFFICE 97  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/307,973A  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/085,190  
 FILING DATE: JUNE 30, 1998  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: FLOYD, LINDA AXAMETHY  
 REGISTRATION NUMBER: 33,692  
 REFERENCE/DOCKET NUMBER: CL-1245  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 302-892-8112  
 TELEFAX: 302-773-0164  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1844 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: double  
 TOPOLOGY: linear  
 MOLECULAR TYPE: DNA (genomic)  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 US-09-307-973A-2

Alignment Scores:  
 Pred. No.: 0.000244 Length: 1844  
 Score: 123.00 Matches: 80  
 Percent Similarity: 39.064 Conservative: 61  
 Best Local Similarity: 22.164 Mismatches: 132  
 Query Match: 6.014 Indels: 88  
 DB: 4 Gaps: 17

US-09-889-746-2 (1-400) x US-09-307-973A-2 (1-1844)

QY 16 SerValAlaValThrGlnGlnLeuTYrAlaGlnProAsnIleuSerLeuProThrValGln 35  
 DB 79 ACCCTGTTGTCACCGCAACCGTTTCAGCACCGCGCGCGTT----- 126  
 QY 36 LeuGlnProValValIleThrIleAspIleSerGlyMetAlaLeuAlaAsnArgIleThr 55  
 DB 127 CTGGCCCGCGTTACATC----- 144  
 QY 56 GlnMetProIleThrThrIleValIleTYrGlnGlnGlnIleGlnGlnAlaIleThrGly 75  
 DB 145 -----GTGACCGCGTCAGGATATGACGCTGGCAA----- 174  
 QY 76 SerArgGlnLeuAlaAspValMetAlaGlnLeu-----IleProSerLeuGly 91  
 DB 175 TCACCTCCGTAATATGTTCTGCGCGGTTCTGGCGTCAATATGCGCAGACGCGC 234

QY 92 ValSerSerGlyThrThrSerAsnProIleGlnThrMetIleGlyArgGlnValGlnPhe 111  
 DB 235 GCGCGCGCAAAATCTCCATTCATTCGCGGACCAACATCGCATGACTGTA 294  
 QY 112 LeuLeuAsnGlyValProLeu-----ThrGlySerArgAspIleSerArg 126  
 DB 295 TTGATTGACGCGCGTGGCTGAATTAGACGGCGTGAGCGGCGCCGATCCACCGAG 354  
 QY 127 GlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyAlaThr 146  
 DB 355 TTCCCGGTGCTGCTG-----GTACACCGATTGAATATATGACGGATCCGCC 402  
 QY 147 Ser---IleTYrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrIleSerAsp 165  
 DB 403 TCCGCTATTATGCTCCGATGCTATCGCGGCGGTGATGATATCATTTGACGCGCGAT 462  
 QY 166 LeuGlnGlnGlnIleGlnIleThrArgIleGlyValIleGlySerIleSerSerGly 185  
 DB 463 AACCA-----GGCAGCAATTAACCGCT--- 486  
 QY 186 GlyIleGly-----TYrGlnValGlyGlnSerValAlaGlyValSerGlyuAsnGly 202  
 DB 487 GGAATGCGAAAGCATAGCTACAG---AATTACGACATCTCGACGCAACGAACTTGGC 543  
 QY 203 AsnValLeu---AlaArgLeuAspValAspTYrArgThrThrGlyGlyAlaPheAspAla 221  
 DB 544 GAATACACGCGCGCGATTCGCGCATTCGCGATTCGATACCAACGAAAGG---TTTGACGTC 600  
 QY 222 AsnGlyLyArgGlyLeuAlaProGluProAlaGlnThrAspIleGlnAspSerIleu 241  
 DB 601 GTACGCAAAAGCGCGTACCGGATGACGAGCGCAGCGGACCGGCGCTTTTGAAGTAA 660  
 QY 242 SerValAsnThrAsnValAspTYrProGlnLeuAspAspIleAsnIleAsnLeuAlaLeu 261  
 DB 661 ACCTTTATGCGCGCGTAAAGCATACCTTTTCTGATCGCGGACGGAATTCGGCTGGT 720  
 QY 262 ThrIleTYrAsnAspIleGlnAspThrAsp---TYrAlaProAspTYrGlyAsnArg 279  
 DB 721 TATGCTAGATTAACGCTACCGCATTCAGACGCGCTATTAATCTCGCG--- 765  
 QY 280 LeuAlaValLeuPheGlyGluIleProSerLeuAsnAlaIleIleGlyLeuSerLeuSer 299  
 DB 766 -----GGCTCGCGCGTGAAT 780  
 QY 300 GlnGlnProIleThrThrIleSerThrPheAsnIleAsnTYrIleIleAspAspLeuTrp 319  
 DB 781 GATACACGCAAACTTATAGCAAAAGCTGAGCGCGGCTGCATTTAAT----- 831  
 QY 320 GlyAsnThrIleAsnThr-----AsnAlaTYrTYrArgArgGluIleGlyArgPhe 336  
 DB 832 GCGGAAGATTCAGCTCAGCTGCTTCAAGCTATAGCCACAGTAAAGATTCAACTAT 891  
 QY 337 TYrProPheValAlaProPheSerIleAlaIleValLeuProIleGlnIleSerMetAsn 356  
 DB 892 GATCCGCACTATGCGCGGATGATACCTCCGCCACGCTGATGATGAACAGTCAAT 951  
 QY 357 Leu 357  
 DB 952 GTT 954  
 RESULT 41  
 US-08-537-361B-3  
 Sequence 3, Application US/08537361E  
 Patent No. 6121037  
 GENERAL INFORMATION:  
 APPLICANT: SciJillKovic, Igor  
 APPLICANT: So, Magdalene  
 APPLICANT: Hwa, Vivian  
 APPLICANT: Hefron, Fred  
 APPLICANT: Nassif, Xavier  
 TITLE OF INVENTION: No. 6121037e1 Bacterial Hemoglobin Receptor  
 TITLE OF INVENTION: Genes and Uses  
 NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: McDonnell Boenhen Hulbert & Berghoff  
 ; STREET: 300 South Wacker Drive, 32nd Floor  
 ; CITY: Chicago  
 ; STATE: Illinois  
 ; COUNTRY: USA  
 ; ZIP: 60606  
 ;  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA: US/08/537,361E  
 ; APPLICATION NUMBER: US/08/537,361E  
 ; FILING DATE: 02-OCT-1995  
 ; CLASSIFICATION: 536  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: No. 6121037nan, Kevin E  
 ; REGISTRATION NUMBER: 35,303  
 ; REFERENCE/DOCKET NUMBER: 94,784-A  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 312-913-0001  
 ; TELEFAX: 312-913-0002  
 ; INFORMATION FOR SEQ ID NO. 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2376 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 1..2373  
 ; US-08-537-361E-3

Alignment Scores:  
 Pred. No.: 2376  
 Score: 122.00  
 Percent Similarity: 36.61%  
 Best Local Similarity: 22.85%  
 Query Match: 5.96%  
 Indels: 126  
 Gaps: 21

US-09-889-746-2 (1-400) x US-08-537-361E-3 (1-2376)  
 Qy 10 LeuSerLeuProLeuSerValAlaValThrGlnGlnLeuThrAlaGlnPro----- 27  
 Db 10 TTACAAATGCCCTATCGCC---GGCTGCTCGGCAGTATTTTCGGCAATCGGCTCTTT 66  
 Qy 28 -----AenGluSerLeuProThr---ValGluLeuGluProValVal 40  
 Db 67 GCGGCAGATGAAGTCGAACCTGAACCCACACCCGTTAAGGCAGAGGTAAGACGAGTGC 126  
 Qy 41 IleThrIleAspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThr 60  
 Db 127 GTTAAAGGTCAGCGCAATCGCTGCGGCTGTGGAACCGGTC-----AAC 171  
 Qy 61 ThrLysValIleThrGluGluGlnIleGlnGlnAlaThrGlySerArgGlnLeuAla 80  
 Db 172 CTTAACCGGTATCAACAGAAATGATACCGCAATAAAGACTGTGGTGCCTATTCCACC 231  
 Qy 81 AppValMetAlaGlnLeuIleProSerLeuGlyValSerSerGlyThrThrSerAsnPhe 100  
 Db 232 GATGTC-----GGCTTACGACGAGGAGCGGTATCAATAA 267  
 Qy 101 Gly-----GlnThrMetHisGlyArgGlnValGlnPheLeuAsnGlyValPro 117  
 Db 268 GCGTTTGGCATTTCGGCGCGTGAAGCGCACCGTGTGCGGCTTAGTATTGACGCGTAAC 327  
 Qy 118 LeuThrGlySerArgAspIleSer-----ArgGlnLeuAsn----- 129  
 Db 328 CTGCCTGATTCCGAGAAACTCGCTGTAGCCCGCTATGGCACTTCAACAGCTCGCT 387

Qy 130 ---SerIleAsnProAsnGlnValAlaAargIleGluValLeuSerGlyAlaThrSerile 148  
 Db 388 CTGCTATCGACCCCGAACTCGTGCACATCGCATCTGTAAGAGGCGGACTCTTC 447  
 Qy 149 ---TyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLysSerAsp----- 165  
 Db 448 AATACCGGCAGCGCGCTTGGCGCGGTGTGAATTACCAACCCCTCAAGGAGCGTAC 507  
 Qy 166 -----LeuGluGluGluGlnPhe-----GluThrArg--- 174  
 Db 508 TTACTGTTTCTGAACGCGAGTTCGGCGTGTATGATAAAACGGTTACAGCAGCGTAAC 567  
 Qy 175 -----IleGlyValHisGlySerLys----- 181  
 Db 568 CGTGAATGACAAATACCTCGTTTCGGCGTGTGACACACCGCGTGTATCGCCTTTG 627  
 Qy 182 LeuSerSerGluGlyIleGlyTyrGlnVal-----GlyGlnSerVal 195  
 Db 628 CTGTATTCGCAACGCGCGCATGAACTGAAGCGCGGCGGAGCGTGTATCCGGTA 687  
 Qy 196 AlaGlyValSerGluAsnGlyAsnValLeuAlaAargLeuAspValAspTyrArgThrThr 215  
 Db 688 GAGGCTGTGTAGCGGAGCGAATATC-----CGT 717  
 Qy 216 GlyGlyAlaPheAspAlaAsnGlyLysArgIleAlaProGluProAlaGlnThrAspLys 235  
 Db 718 GGTTCGCG-----CGCGTATTCGTATCGCTCCCA----- 750  
 Qy 236 GlnAspSerLysSerLeuSerValAsnThrAsnValAspTyrGlnLeuAspAspLysGln 255  
 Db 751 -----CACAAATACCAAGCTTCTGGGTAGATGTTGTATCAANTCAACGACCAAC 804  
 Qy 256 AsnIleAsnLeuAlaLeuThrHisTyrAsnAspLysGlnAspThrAspTyrAlaProAsp 275  
 Db 805 CGCATCGCGCATCGCTC-----AACGTCAGCAGCGGCATAATTACACGTTGAA 855  
 Qy 276 TyrGlyAsnArgLeu-AlaValLeuPheGlyGlyLysProSerLeuAsnAlaIle----- 293  
 Db 856 GAGTCTTACAACTCTGTTCTTATTGGCGTGAAGCTGACGATGTCAACAGACGCGGT 915  
 Qy 294 -----LysGln 295  
 Db 916 AACACCAACTCTTTTACGAATGAGCGCGGAATCCGACCGGTGTCTATGTTAAAGCG 975  
 Qy 295 yLeuSerLeuSerGluGlnProLys-----ThrThrLysSerThrPheAsnIle 311  
 Db 976 GATGTCGATTATCAAAACCAACCAAGTATCTCGGTCAACTACAAAGGT-TCGTTCCCGAC 1034  
 Qy 311 eAsnTyrHisAspAspLeuTyrGlyAsnThrIleAsnThrAsnAlaTyrTyrArgAr 331  
 Db 1035 GAATTAC-----ACCACATGGGAACCGAGTACCATATAAAA 1070  
 Qy 331 gGluLysGlyArgPheTyr 337  
 Db 1071 GGAAGTTGCGGAATCTAT 1089  
 RESULT 42  
 US-08-817-707-3  
 ; Sequence 3, Application US/08817707  
 ; Patent No. 6277382  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Stojiljkovic, Igor  
 ; APPLICANT: So, Magdalene  
 ; APPLICANT: Hwa, Vivian  
 ; APPLICANT: Heffron, Fred  
 ; APPLICANT: Nassif, Xavier  
 ; TITLE OF INVENTION: No. 6277382el Bacterial Hemoglobin Receptor  
 ; TITLE OF INVENTION: Genes and Uses  
 ; NUMBER OF SEQUENCES: 15  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: McDonnell Boenhen Hulbert & Berghoff  
 ; STREET: 300 South Wacker Drive, 32nd Floor  
 ; CITY: Chicago

```

STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/817,707
FILING DATE: 19-AUG-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: No. 6277382nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 94,784-J
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-913-0001
TELEFAX: 312-913-0002
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2376 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2373
US-08-817-707-3

```

```

Alignment Scores:
Pred. No.: 0.00483 Length: 2376
Score: 122.00 Matches: 93
Percent Similarity: 36.61% Conservative: 56
Best Local Similarity: 22.85% Mismatches: 133
Query Match: 5.96% Indels: 126
DB: 4 Gaps: 21

```

US-09-889-746-2 (1-400) x US-08-817-707-3 (1-2376)

```

QY 10 LeuSerLeuProLeuLeuSerValAlaValThrGlnGlnLeuTyraAlaGlnPro----- 27
DB 10 TTAACAATGCGCCCTATCGCC---GCGCTGCGGAGTATTTTCGCAATCGGCTTTT 66
QY 28 -----AenGlnSerLeuProThr---ValGlnLeuGlnProValVal 40
DB 67 GCGGCAAGTGAAGCTCAACTGAACCAACACCGTTAAGGCAAGGTAAAGACAGTGGCG 126
QY 41 IleThrIleAspIleSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHISThr 60
DB 127 GTTAAAGTCAGGCAATGCGCTGCGCTGTGTGAAGCGCTC-----AAC 171
QY 61 ThrIleValIleTyrgluGlnGlnIleGlnGlnAlaThrIleSerArgGlnLeuAla 80
DB 172 CTTAACCGTATCAAAAGAAATGATACGCAATTAAGACTTGGTGGCTTATTCACAC 231
QY 81 AspValMetAlaGlnLeuIleProSerLeuGlyValSerSerGlyThrThrSerAsnIle 100
DB 232 GATGTC-----GGCTTGAGCAACGAGGCCCTCATCAAAA 267
QY 101 Gly-----GlnThrMetHISGlyArgGlnValGlnPheLeuLeuAsnGlyValPro 117
DB 268 GCGTTTGCCATTGCGCGCGTGAAGCGACCGTGTGCGCTTGTATGTAACGCGTAAAC 327
QY 118 LeuThrGlySerArgAspIleSer-----ArgGlnLeuAsn----- 129
DB 328 CTGCTGATTCGCAAGAAACTCGCTGTACGCGCTTATGCAACTTCAACAGCTGCGGT 387
QY 130 ---SerIleAsnProAsnGlnValAlaArgGlnGlnValLeuSerGlyValaThrSerIle 148
DB 388 CTGCTATGACCCGCAACTGTCGCGCAACATGACATCGTAAAGGGGCGGACTTTTC 447

```

```

QY 149 ---TyrglySerGlyAlaThrGlyLeuIleAsnIleValThrIleSerAsp----- 165
DB 448 AATACCGGAGCGCGCGCTTGGCGCGGTGTGAATTAACCAACCTGCAAGACGTAC 507
QY 166 -----LeuGlnGlnGlnPhe-----GlnThrArg--- 174
DB 508 TTACTGTTCCTGAACCGGAGTTCGCGGTGATGATGAAAAACGGTTACAGACGCGTAAC 567
QY 175 -----IleGlyValHISGlySerIle----- 181
DB 568 CGTGAATGACAAATACCTCGCTTTCGCGGTGAGCAACGACCGCGTGAATCGCTTGG 627
QY 182 LeuSerSerGlnGlyIleGlyTyrglnVal-----GlyGlnSerVal 195
DB 628 CTGTATTCGCAACGCGCGCGCATGAATACTGAAGCGCGGCAAGCGGTGATTACCGGTA 687
QY 196 AlaGlyValSerGlnAsnGlnValLeuAlaArgLeuAspValAspTyraThrThr 215
DB 688 GAGGTGCTGTGAGCGGACGCAATATC-----CGT 717
QY 216 GlyGlyAlaPheAspAlaAsnGlyValArgIleAlaProGluProAlaGlnThrAspIle 235
DB 718 GGTTCGCG-----CGCGTATTCCTGATCCTCCCA----- 750
QY 236 GlnAspSerIleSerLeuSerValAsnThrAsnValAspTrpGlnLeuAspAspIleGln 255
DB 751 -----CACAAATACCAACGCTTCTGGTAAAGATTGCTTATCAATCAACGCAACGAC 804
QY 256 AsnIleAsnLeuAlaLeuThrIleSTyraAsnAspIleGlnAspThrAspTyraIleAsp 275
DB 805 CGCATCGCGGCAATCGCTC-----AACGCTCAGACAGGCGCATATATCAACGCTTGA 855
QY 276 TyrglyAsnArgLeuAlaValIlePheGlyGlyIleProSerLeuAsnAlaIle----- 293
DB 856 GAGCTTAAACCTGCTCTTCTTATTTGCGGTGAAGCTGACGATCAACAGACGCGCT 915
QY 294 ----- 295
DB 916 AACCAACCTCTTTACGAATGAGCGCGGAATCCGACCGGTGTGCTATGTAAGAGG 975
QY 295 yLeuSerLeuSerGlnGlnProIle-----ThrThrIleSerThrPheAsnIle 311
DB 976 GATGTCGATTAATCAAAAAACCAAGTATCTGCGGTCAACTACAAAGGT-TCGTCCCGAC 1034
QY 311 eAsnTyraHISAspAspLeuTrpGlyAsnThrIleAsnThrAsnAlaTyraArgArg 331
DB 1035 GAATTAAC-----ACCAATGGAACCGAGTACCATTAATAA 1070
QY 331 gGlnyGlyArgPheTyra 337
DB 1071 GGAAGTTGGCGAAATCTAT 1089

```

```

RESULT 43
US-08-817-707-5
; Sequence 5, Application US/08817707
; Patent No. 6277382
GENERAL INFORMATION:
APPLICANT: Stojiljkovic, Igor
APPLICANT: So, Magdalene
APPLICANT: Hwa, Vivian
APPLICANT: Heffron, Fred
APPLICANT: Nasif, Xavier
TITLE OF INVENTION: No. 6277382el Bacterial Hemoglobin Receptor
TITLE OF INVENTION: Genes and Uses
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSER: McDowell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive, 32nd Floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/817,707  
FILING DATE: 19-AUG-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Mr. 6277382han, Kevin E  
REGISTRATION NUMBER: 35,303  
REFERENCE/DOCKET NUMBER: 94,784-J  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-913-0001  
TELEFAX: 312-913-0002  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2373 base pairs  
TYPE: nucleic acid  
STRAINEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..2370  
US-08-817-707-5



```

CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/124,254
FILING DATE: 20-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/973,336
FILING DATE: 05-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/572,187
FILING DATE: 23-AUG-1990
ATTORNEY/AGENT INFORMATION:
NAME: Gallagher, Thomas C.
REGISTRATION NUMBER: 37,066
REFERENCE/DOCKET NUMBER: SPA-1-PDC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 645-1405
TELEFAX: (212) 645-2054
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3537 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORGANISM: Neisseria meningitidis
STRAIN: FAM18, FAM20, B18b6, group X and group W135
FEATURE:
NAME/KEY: CDS
LOCATION: 721..3450
FEATURE:
NAME/KEY: mat-peptide
LOCATION: 793..3447
US-08-363-124A-3

```

```

Alignment Scores:
Pred. No.: 0.00121 Length: 3537
Score: 121.00 Matches: 89
Percent Similarity: 34.49% Conservative: 60
Best Local Similarity: 20.60% Mismatches: 161
Query Match: 5.91% Indels: 122
DB: 2 Gaps: 23

```

```

US-09-889-746-2 (1-400) X US-08-363-124A-3 (1-3537)
QY 1 MetArgHisSerHisrPhgIntPLeuSerLeuProLeuSerValAlaValThr 20
DB 721 ATCAACAGCAACATTTGTTCCGA---TTAAATATTTATGCTGTTTAAAGACCGCG 777
QY 21 GlnGlnLeuTyAlaGln-----ProAenGluSerLeuProThrValGluLeuGluPro 38
DB 778 CTGCCCGTTATTCAGAAATGTGCAAGCCGCAACAGCACAGAAAAACAGTTGATACC 837
QY 39 ValValIleThrIleAspIysSerGlyMetAlaLeuAlaSerGlyIleThrGlnMetPro 58
DB 838 ATACAGGTAAAGCCAAAAACAGAAAAACCCCGCGGATACAGAAAGTAACAGGCGCTGGGC 897
QY 59 HisThrThrIleValIleTyrgIuGluGlnIleGlnIuGlnAlaThrGlySerArgGln 78
DB 898 AAGTGTGTCAAG---TCTTCGATAGCTTAAGTAAGAAACAGGTTTGAATATCCAGAC 954
QY 79 LeuAlaSerValMetAlaGlnLeuIleProSerLeuGly--ValSerSerGlyThrThr 97
DB 955 CTG-----ACCCCTTATGATCCGGATTCGCGTGGCGAAACAGGCGTGGCGGC 1002
QY 98 SerAspPheGlyGlnThrMetHisGly-----ArgGlnValGlnPheLeuLeuAsn 114
DB 1003 GCAAAGTTCGGCTATTCAATACGCGGATGATGATAAAACCGCGTTCTTAAACGATGAC 1062
QY 115 GlyVal-----ProLeuThrGlySerArgAspIle 124

```

```

DB 1063 GCGGTTTCCGAATACAGTCTTACACCGCGGAGCGGCGCATTTGGTGGAGCGAGCGCG 1122
QY 125 -----SerArgGlnLeuAsnSerIleAsnProAenGlnValAlaValIleGluValLeu 142
DB 1123 GGTAGCAGCGCGCGCATCATGAAATTCAGATGAAACGTAAGCCCTTGAATACACC 1182
QY 143 SerGlyAla---ThrSerIleTyrgIysSerGlyAlaThrGlyGlyLeuIleAsnIleVal 161
DB 1183 AAGGTTCAATTCATCAAAATACGAAACGCGGATTCGAGGTTCCGATTCGATTCAA 1242
QY 162 ThrIle-----SerAspLeu-----GluGluGluGlnPheGluThrArgIleGlyVal 177
DB 1243 ACCAAACCGCAGCCGACATATTCGAGAGGAGGAAAAACAGTGG-----GGCATTT 1290
QY 178 HisGlySerIleLeuSerSerGlyIleGlyTyrgIuValGlyGlnSerValAlaGly 197
DB 1291 CAG---AGTAAACCTGCTATTTCGGAAAAAGACCATGCCCTGACCATCCCTGGCTT 1347
QY 198 ValSerGluAsnGlyAsnValLeuAlaArg----- 207
DB 1348 GCGGACGAGCGCGCGCGCGGAGGCCCTTATTATTAATAACGCGGGGTCCGGA 1407
QY 208 -----LeuAsp 209
DB 1408 ATCCATGCGCATTAAGATCCCGCAAGGGGTGACAGCTTCAACCGGCTGTGTGAC 1467
QY 210 ValAspTyArgThrThrGlyValAlaPheAspAla----- 221
DB 1468 GAGGACAAAGAGAGAGGTGGCAGTCAATGATATTCATTTGTCGAGAGAGAAATGCCAC 1527
QY 222 AenGlyTyArgIleAlaProGluProAlaGlnThrAsp-----LysGlnAspSer 238
DB 1528 AATGATATGCGCGCTGTAAAAACAGCTGAAAGAAATGCTCGTCAAAATGAGCGC 1587
QY 239 LysSerLeuSerValAsnThr----- 245
DB 1588 AAACCGTCAGCAGCAGGATTATACCGGCTCCAAACCGCTTACTTCCGAACCGCTTGAG 1647
QY 246 -----AsnValAspTyrgIleAspIysSerGlnAsnIle 257
DB 1648 TATGCGACCAATCATGCTGTTCGACCGGGTGGCATTTGACAAACCGCATTAATGTC 1707
QY 258 AsnLeuAlaLeuThrHisTyArgAsnAspIysGlnAspThr----- 270
DB 1708 GAGCGCTTCTCAAGCTACGACGACGACCTTGTATACCGGATATGACTGTTCTGCC 1767
QY 271 -----AspTyAlaPro-----AspTy 276
DB 1768 TATTTACAGAGAAATATGATGACCGGTTGCTGAAAGTCTTGGCAATATTCGGGC 1827
QY 277 GlyAsnArgLeuAlaValLeuPhe-----GlyGluLysProSerLeuAsnAlaIle--- 293
DB 1828 GATTAATAGCAGAAAGGCTGTTTGTTCAGGAGAGGCGCATTTGACGAGGTACCGGT 1887
QY 294 LysGlyLeuSerLeuSerGlyGlnProLysThrThrIleSerThrPheAsnIleAsnTy 313
DB 1888 TACGTTACCGGCGCTTTTATGATGACCCCATCTAAAAACCGCTACCGGGGTGCAATAT 1947
QY 314 ---HisIleAsp-----AspLeuTyrgIyAsn 321
DB 1948 GTTATACATATATGCTGATAAGATACCTGGCCGAT 1983

```

```

RESULT 45
US-08-448-194-3
Sequence 3, Application US/08448194
Patent No. 6028049

```

```

GENERAL INFORMATION:
APPLICANT: JACOBS, Eric
APPLICANT: LEGRAIN, Michele
APPLICANT: MAZARIN, Veronique
APPLICANT: BOUCHON-THRISEN, Bernadette
APPLICANT: SCHRYVERS, Anthony B.

```

APPLICANT: BLOCH, Marie-Aline  
TITLE OF INVENTION: DNA FRAGMENTS CODING FOR THE TRANSFERRIN  
RECEPTOR OF NEISSERIA MENINGITIDIS  
NUMBER OF SEQUENCES: 62  
CORRESPONDENCE ADDRESS:  
ADDRESS: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/448,194  
FILING DATE: 23-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

```

Db 1147 GATAATTAAGCGAAGGCTGTTGTTCAAGGAGAGGCGAGTACATTGCAGGGTATCGGT 1206
Oy 294 LysGlyLeuSerLeuSerGluGlnProLySerThrLySerThrPheAsnIleAsnTyr 313
    |||
    ::|||
Db 1207 TACGGTACCGGCGGTGTTTATGATGAACCCATCTAATAAACCGCTACGGGTCGAATAT 1266
Oy 314 ---HisHisAsp-----AspLeuTyrGlyAsn 321
    ::|||
    ::|||
Db 1267 GTTACCATATATGCTGATAAGGATACCTGGCCGAT 1302

```

Search completed: December 25, 2002, 22:55:26  
 Job time : 188 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

# OM protein - nucleic search, using frame\_plus\_p2n model

Run on: December 25, 2002, 20:33:54 ; Search time 1957 Seconds

(without alignments)  
3310.268 Million cell updates/sec

Title: US-09-889-746-2

Perfect score: 2047

Sequence: 1 MRSHYFQWLSLPLSVAVT.....SKAVLGRVPLNKRALF 400

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgm2.1/USPTO\_spool/US09888746/runat.20122002.143749.14216/app.query.fasta.1.583  
-DB=EST -QFMT=fastag -SUFFIX=1st -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -STRAT=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-LOCALALIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=45 -MODE=LOCAL  
-OUTFMT=PCO -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09888746.ccdn.1.1.899 @runat.20122002.143749.14216 -NCPU=6 -ICPU=3  
-NARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estmu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hcc.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hcc.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: gb\_gss.\*  
18: em\_gss\_hum.\*  
19: em\_gss\_inv.\*  
20: em\_gss\_pin.\*  
21: em\_gss\_vrt.\*  
22: em\_gss\_fun.\*  
23: em\_gss\_mam.\*  
24: em\_gss\_mus.\*  
25: em\_gss\_othet.\*  
26: em\_gss\_pro.\*  
27: em\_gss\_tod.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	376.5	18.4	573	17	BH377354
2	338.5	16.5	443	17	BH391113
3	284.5	13.9	655	17	BH379141
4	184	9.0	766	17	BH388728
5	140.5	6.9	598	17	BH380398
6	125	6.1	530	12	BG023387
7	120	5.6	519	17	AZ302660
8	114	5.9	538	17	BH397318
9	113	5.5	558	17	BH382603
10	112.5	5.5	428	17	BH001150
11	112	5.5	423	17	P132R
12	106	5.2	683	17	BH825061
13	106	5.2	727	9	AJ452021
14	105	5.1	3064	11	AK004607
15	104.5	5.1	694	17	AQ989725
16	102	5.0	839	13	B1158313
17	102	5.0	1344	11	BC017847
18	102	5.0	6042	11	BC010844
19	101.5	5.0	634	17	BH368633
20	101.5	5.0	638	13	BM358704
21	101.5	5.0	1240	11	A1105884
22	99.5	4.9	856	17	A2533394
23	98.5	4.8	791	14	B0682892
24	98	4.8	526	17	CNS00W7R
25	98	4.8	593	14	B0837115
26	97	4.7	615	13	B0035098
27	97	4.7	812	14	B0748998
28	97	4.7	841	17	CNS070462
29	97	4.7	1103	14	B0712865
30	97	4.7	917	13	BM461330
31	97	4.7	1482	11	AK011327
32	96.5	4.7	605	14	B0835748
33	96.5	4.7	605	14	B0836176
34	96.5	4.7	793	17	BH378768
35	96.5	4.7	973	14	B0712258
36	96.5	4.7	1005	17	BH770394
37	96	4.7	509	10	BE015300
38	96	4.7	593	14	B0836774
39	96	4.7	749	9	A1133960
40	95.5	4.7	863	17	BH396248
41	95	4.6	590	17	BH370350
42	95	4.6	979	17	CNS06PDW
43	94.5	4.6	605	14	B0836244
44	94.5	4.6	611	14	B0835870
45	94.5	4.6	649	13	B0423365

## ALIGNMENTS

RESULT 1  
BH377354/c  
LOCUS BH377354  
DEFINITION AG-ND-13701.TR ND-TM Anopheles gambiae genomic clone AG-ND-13701,  
DNA sequence.  
ACCESSION BH377354  
VERSION BH377354.1 GI:17323496  
KEYWORDS  
SOURCE African malaria mosquito.  
ORGANISM Anopheles gambiae  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
Anophelinae.  
REFERENCE  
1 (bases 1 to 573)  
Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.

TITLE Direct Submission of BAC-end sequences from Anopheles gambiae  
JOURNAL Unpublished (2001)  
COMMENT Other\_GSSs: AG-ND-13701.TF  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: b.loftus@igr.org  
This clone is from an A. gambiae BAC library (ND-TAM) provided by  
F.H. Collins and sequenced by The Institute for Genomic Research  
(TIGR). The BAC library was generated from A. gambiae PEST strain  
DNA. All DNA was extracted from newly hatched first instar larvae  
to minimize the inclusion of DNA from microorganisms that inhabit  
the gut. The DNA is derived from mixed sexes of larvae. The BAC  
library was constructed at Texas A&M University BAC Center  
University, College Station, Texas 77843-2123, USA using a HindIII  
partial digest.  
Seq primer: M13 Rev  
Class: BAC ends.

FEATURES  
source  
1..573  
/organism="Anopheles gambiae"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-13701"  
/clone\_lib="ND-TAM"  
/note="Vector: pECBAC1; Site\_1: HindIII"

BASE COUNT 145 a 137 c 111 g 180 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.26e-33 Length: 573  
Score: 376.50 Matches: 84  
Percent Similarity: 63.16% Conservative: 36  
Best Local Similarity: 48.21% Mismatches: 56  
Query Match: 18.39% Indels: 14  
DB: 17 Gaps: 4

US-09-889-746-2 (1-400) x BH377354 (1-573)

Qy 10 LeuSerLeuProLeuLeuSerValAlaValThrGlnGlnLeuThrAlaGln----- 26  
Db 557 ATCAACTACGCTGTTAAGTTGGCTATTGGACACAGATTATGCGCAAGCCACG 498  
Qy 27 -----ProAsnGluSerLeuProThrValGluLeuGluProValValIleThr 42  
Db 497 ACTCAAAAAGTTGAATATCTGGGCGAGAGGTAAAGGCTAGCGCAATCGTGTCA 438  
Qy 43 IleAspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrLys 62  
Db 437 GCAACACGTTCT-----GCAAAAAGTATTTCAGACATTTCAGCGGCGGTAT 390  
Qy 63 ValIleThrGluGluGlnIleGlnGlnAlaThrGlySerArgGlnLeuAlaAspVal 82  
Db 389 CGCATTGGAACGTAATGAAATTGAAAACAAGCCAAATCGCGGTAAAGCAGCGCAAT 330  
Qy 83 MetAlaGlnLeuLeuProSerLeuGlyValSerSerGlyThrThrSerAsnPheGlyGln 102  
Db 329 TTAGGGCTATTGTTCTTCACTTCAGCCCAATACGTGTACAACTTCTATTATGTTATG 270  
Qy 103 ThrMetHisGlyArgGlnValGlnPheLeu-LeuAsnGlyValProLeuThrGlySerAr 122  
Db 269 ACCATGCGCGCGCGTGTGTGCAATATATGATTGGATGGATACCAACAAGGCTCAG 210  
Qy 122 GAspIleSerArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGluVal 142  
Db 209 GGATGGATCGCGTCACCTCAATAGCATTTCTTCTCAATGATGGACAGAAATAGAGTCT 150  
Qy 142 uSerGlyAlaThrSerIleThrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValTh 162  
Db 149 GTCTGGGGCAACGAGTATTATTGGCTCAGGTGCACTGCGGCTATTATCAATATCAT 90

Qy 162 rIysSerAspLeuGluGluGlu---GlnPheGluThrArgIleGlyVal---HisGlySe 180  
Db 89 CAAAAAGCGCGTGAAGCTTTAAGTTTGAACATAACTGGGCTAACCTCTGGGAA 30  
Qy 180 rIysLeuSerSerGluGlyIleGlyTyr 189  
Db 29 TAAATTTAAAGCGGATGCTTGCCTTAT 2

# RESULT 2

BH391113/c BH391113 443 bp DNA linear GSS 11-DEC-2001  
LOCUS AG-ND-13701.TR.1 ND-TAM Anopheles gambiae genomic clone AG-ND-13701  
DEFINITION , DNA sequence.

ACCESSION BH391113

VERSION BH391113.1 GI:17337254

KEYWORDS GSS.

SOURCE African malaria mosquito.

ORGANISM Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;

Anopheles.

REFERENCE 1 (bases 1 to 443)

AUTHORS Shetty, J., Malek, J., Koo, H., Collins, F., Gardner, M. and Loftus, B.J.

TITLE Direct Submission of BAC-end sequences from Anopheles gambiae

JOURNAL Unpublished (2001)

COMMENT Other\_GSSs: AG-ND-13701.TF.1

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: b.loftus@igr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by  
F.H. Collins and sequenced by The Institute for Genomic Research  
(TIGR). The BAC library was generated from A. gambiae PEST strain  
DNA. All DNA was extracted from newly hatched first instar larvae  
to minimize the inclusion of DNA from microorganisms that inhabit  
the gut. The DNA is derived from mixed sexes of larvae. The BAC  
library was constructed at Texas A&M University BAC Center  
University, College Station, Texas 77843-2123, USA using a HindIII  
partial digest.  
Seq primer: M13 Rev  
Class: BAC ends.

FEATURES  
source  
1..443  
/organism="Anopheles gambiae"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-13701"  
/clone\_lib="ND-TAM"  
/note="Vector: pECBAC1; Site\_1: HindIII"

BASE COUNT 115 a 107 c 87 g 134 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 2.38e-29 Length: 443  
Score: 338.50 Matches: 68  
Percent Similarity: 70.80% Conservative: 29  
Best Local Similarity: 49.84% Mismatches: 35  
Query Match: 16.54% Indels: 5  
DB: 17 Gaps: 3

US-09-889-746-2 (1-400) x BH391113 (1-443)

Qy 59 HisThrThrLysValIleTyr-----GluGluGlnIleGlnGlnAlaThrGly 75  
Db 412 CACTTCACAGCGCGCGTGTATGCGATTGAACGTAAATGAAATTAACCAACCCATCG 353  
Qy 76 SerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSerGly 95  
Db 352 GGTAAAGCACGCGCAGAAATTTAGGGCTATTGGTCTTCTTCACTCAGCCCAATCTG 293

Oy		96	ThrtTsrAsnPhgClGlnThrncHtsglYvrgGlInPheLeuAnngly	115
Dd		232	ACAACTTTAATTATGGTAAGTCACCATGCCGCCTGTGTGTCATAATATATGAATGAGG	233
Oy		116	VAlProLeuthrGIyserrAgapileserragInLeuanSerlleasnbPrOansgnin	135
Dd		232	ATACCACAACAGCGCTCACGGGATGGATGGCTCAGCTCAATAGCATTTCTCCCTCAANG	173
Oy		136	vAlaaargllleglvaleusertglYalatrserllerTyglySertglYalathrgly	155
Dd		172	ATTGACAGAATVGAAGTGCTGTGGGGCAACAGACTTATATGCTCAGAGTCCAAGTgc	113
Oy		156	GlyleulleasnilleValthrIryserrApelaeplugluGu---GlnpheguTrarag	174
Dd		112	GGTATATCAATATCATTRCCAAAAAGCGCGTAGTGMAGCTTMAAGTTTTGAACATAA	53
Oy		175	Ileglvval--HisglySerlysleusenSergluglyIlleglyTrygin	190
Dd		52	CTGGGGCTAACCTCTGGGAATATTTAAAAGCGATGCTGGCCTTATGAA	2
RESULT 3 BH37914L/c		BH379141	655 bp DNA linear GSS 10-DEC-2001	
LOCUS		AG-ND-139D9.TF ND-TAM Anopheles gambiae genomic clone AG-ND-139D9,		
DEFINITION		DNA sequence.		
VERSION		BH379141		
KEYWORDS		BH379141.1 GI:17325283		
SOURCE		GSS:		
ORGANISM		African malaria mosquito. <i>Anopheles gambiae</i> Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoides;		
REFERENCE		1 (bases 1 to 655) Shetty,J.J., Malek,J.S., Koo,H., Collins,F., Gardner,M. and Loftus,B.U		
AUTHORS		Title Direct Submission of BAC-end sequences from <i>Anopheles gambiae</i>		
JOURNAL		Unpublished (2001)		
COMMENT		Contact: Brendan J Loftus Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel.: 301 838 0208 Fax: 301 838 3543 Email: b.loftus@igr.org This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PBST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest. Seq primer: M13 For Class: BAC ends. Location/Qualifiers . . . 655 /organism="anopheles gambiae" /strain="PBST" /db_xref="taxon:7165" /clone="AG-ND-139D9" /clone_1fb="ND-TAM" /note="Vector: pECBACL; Site_1: HindIII"		
FEATURES		source		
BASE COUNT		156 a 153 c 120 g 226 t		
ORIGIN				
Alignment Scores: Pred. No.: Score: Percent Similarity: Best Local Similarity: Query Match:		9..91e-23 284.50 55.16% 33.63% 13.90%	Length: Matches: Conservative: Mismatches: Indels:	655 75 48 80 20

[illegible]

9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjloftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.  
Seq primer: M13 For  
Class: BAC ends.

# FEATURES

source  
1. .766

Location/Qualifiers  
/organism="Anopheles gambiae"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-1045"  
/clone\_lib="ND-TAM"  
/note="Vector: pECBAC1; Site 1: HindIII"

BASE COUNT 240 a 145 c 108 g 273 t

Alignment Scores:  
Pred. No.: 8,05e-11 Length: 766  
Score: 184.00 Matches: 49  
Percent Similarity: 56.00% Conservative: 35  
Best Local Similarity: 32.67% Mismatches: 54  
Query Match: 8.99% Indels: 12  
DB: 17 Gaps: 4

US-09-889-746-2 (1-400) x BH388728 (1-766)

QY 16 SerValAlaValThrGlnGlnLeuTyAlaGlnProAsnGluSerLeuProThrValGlu 35  
DB 445 TCTATAGCTGTGGTCTCTACACACATCTGAGAAGAA-----GAG 401  
QY 36 LeuGluProValValThrLeuThrLeuAspSerGlyMetAlaLeuAlaAsnArgIleThr 55  
DB 400 ATTGAAGAAGTTGTGTA-----TCGGAAATACTTTCGAGCAGAAAGTAAAG 353  
QY 56 GlnMetProHisThrThrIleValIleTyGluGluGlnIleGlnGlnAlaThrGly 75  
DB 352 GAAGTTCCTTATTCCTATTAAGGTTATAGATAAATAACACAGATTTCAGCAG-----TCAGGA 299  
QY 76 SerArgGlnLeuAlaAspValMetAlaGlnLeuLeuProSerLeuGlyValSerGly 95  
DB 298 AGTGACGATTAGTGATATTCTTCGAGCAGCAACGGTCTTATTATTACCCAAACCAT 239  
QY 96 ThrThrSerAsnPheGlyGlnThrMetHisGlyArgGlnValGlnPheLeuLeuAsnGly 115  
DB 238 GGTACAAGCTCGAATGCGAGGATGACCGGTGAATATACATACTGATAATGGA 179  
QY 116 ValProLeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAsnProAsnGln 135  
DB 178 GAACCTTCGTCGGAAGGACTCGGGAACACTGGATCTTTCAGAAATACCCGTAATAT 119  
QY 136 ValAlaArgIleGluValLeuSerGly---AlaThrSerIleTyGlySerGlyAlaThr 154  
DB 118 ATTAGAGATATAGAAATTAAGGACCCAGTCTTCTACTATACGATCGGATGCTTTA 59  
QY 155 GlyGlyLeuIleAsnIleValThrIleSer 164  
DB 58 GCCCGAGTGGTAAATATCATCATCCGAAACA 29

RESULT 5  
BH380398/c BH380398 598 bp DNA linear GSS 10-DEC-2001  
LOCUS AG-ND-16114.TR ND-TAM Anopheles gambiae genomic clone AG-ND-16114,  
DEFINITION DNA sequence.

ACCESSION BH380398  
VERSION BH380398.1 GI:17326540  
KEYWORDS GSS.  
SOURCE African malaria mosquito.  
ORGANISM Anopheles gambiae

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.

REFERENCE 1 (bases 1 to 598)  
AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.  
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae  
JOURNAL Unpublished (2001)  
COMMENT Other GSSs: AG-ND-16114.TP

Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjloftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PEST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.  
Seq primer: M13 Rev  
Class: BAC ends.

# FEATURES

Location/Qualifiers  
source  
1. .598

/organism="Anopheles gambiae"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-16114"  
/clone\_lib="ND-TAM"  
/note="Vector: pECBAC1; Site 1: HindIII"

BASE COUNT 164 a 127 c 95 g 212 t

Alignment Scores:  
Pred. No.: 6.84e-06 Length: 598  
Score: 140.50 Matches: 56  
Percent Similarity: 39.73% Conservative: 31  
Best Local Similarity: 25.57% Mismatches: 77  
Query Match: 6.86% Indels: 55  
DB: 17 Gaps: 11

US-09-889-746-2 (1-400) x BH380398 (1-598)

QY 173 ThrArgIleGlyValHisGlySerLysLeuSerSer-----Glu 185  
DB 597 ACTGTCCTGGGATTTACCGGCGCATGAATTATCAATCAAGATGTTTAAATCTGAGGA 538  
QY 186 GlyIleGlyTyGlnValGlyGlnSerValAlaGlyValSerGluAsnGlyAsnValLeu 205  
DB 537 GGAGCCGATATAGATATACAGAGCTTTTATGGA-----AAGCTG 496  
QY 206 AlaArgLeuAspValAspTyArgThrThrGlyAlaPheAspAlaAsnGlyLysArg 225  
DB 495 GCGAAATTTGATTAT-----CTGGTAAACCGAACCTTTCACAGAAATGGTGTAAAG 445  
QY 226 IleAlaProGluPro---AlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsn 244  
DB 444 ATAGATGAGAGAGGTTTAGTCCAAACCCAGATATGAGCTTGGTGAACAATGGTGAAT 385  
QY 245 -----ThrAsnValAspTrpGlnLeuAspAspLysGlnAsnIleAsnLeuAlaLeu 261  
DB 384 AATGCTTTTGCAAAATGGGTACTCTCTCAATGACAAAAGCAGATAGATGATGATAC 325  
QY 262 ThrHisTyAsnAspLysGluAspThrAspTyAlaProAspTyGlyAsnArgLeuAla 281



```

Db 324 AATTACTATTCAGTTTGGAGATTCCTCAATATATATCTGACACAAAGAAAG----- 274
Oy 282 ValLeuPheGluGluProSerLeuAanAlaIle---LyGlyLeuSerLeuSerGlu 300
Db 273 -----TATGGGAAAGGCGCTCACTGTAATATGAGAGGCTCTTGTTAAAGAA 220
Oy 301 GlnProLySerThrLySerThrPheAanIleAanTyHisHisAAspAseUtrpGly 320
Db 219 GGAACATAATATATATACACACAGCTTAT---TTGAATATATGCAATCTTATATCCGT 163
Oy 321 AantThr---IleAanThrAanAlaTyTyTArgArgGluLeuGlyArgPheTyProPhe 339
Db 162 AATACATCCCTCAATCATCCCTTATTTTCA-----GACCTTTAT----- 121
Oy 340 ValAlaProPheSerIleAlaLySaIeAeuProIleLeuGlnSerMetAanLeuProSer 359
Db 121 ----- 121
Oy 360 AlAThrLeuAAspAlaTyTyThrLySaIeProGlnAlaArgAlaTyGlyValLeuGln 378
Db 120 -----ACATTTTAGACTACAGAGAAAGCTCCCGTTGAAACAGGTGCGCAATCTGCA 67

RESULT 6
LOCUS BG023387 530 bp mRNA linear EST 31-MAY-2002
DEFINITION dg40h09.x1 Xenopus laevis gastrula non normalized Xenopus laevis
CDNA clone XENOPUS SOURCE ID:xlmgaa011018 3' similar to TR:Q9XBV1
Q9XBV1 TONB-LINKED RECEPTOR TLR. ; mRNA sequence.
BG023387
VERSION BG023387.1 GI:12479466
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 530)
Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Peterson
,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterston,R. and Wilson,R.
Washu Xenopus EST project, 1999
Unpublished (1999)
Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Library constructed by Bruce Blumberg
DNA Sequencing by: Washington University Genome Sequencing Center
This clone is available royalty-free through LML; contact the
IMACS Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco
High quality sequence stop: 370.
Location/Qualifiers
1. 530
/organism="Xenopus laevis"
/db xref="taxon:8355"
/clone="XENOPUS SOURCE ID:xlmgaa011018"
/clone_1ib="Xenopus laevis gastrula non normalized"
/issue_type="gastrula (stages 10.5, 11.5 mixed)"
/lab_host="Top-10 F"
/note="Vector: pBlueScript SK-; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from 2ug of poly A+ RNA (equal
parts from stage 10.5 and stage 11.5 gastrulae).
EcoRI-XhoI cut cDNA was then ligated into UniZap-XR
(Stratagene) with EcoRI at the 5' end and XhoI at the
3' end. SS-library phagmids were prepared by mass excision
from the original library and normalized by hybridization

```

```

US-09-889-746-2 (1-400) x BG023387 (1-530)
Oy 116 ValProLeuThrGlySerArgAspIleSerArgGlnLeuAanSerIleAanProAanGln 135
Db 467 GTGCTGTGACCGGCGCGCGCTCCGATTCGACACGCTTCACACCGCGCGGAAAGC 408
Oy 136 ValAlaArgIleGluValLeuSerGly---AlaThrSerIleTyGlySerGlyAlaThr 154
Db 407 ATCGCCCGCGTGAAGTGTGCTGGCGCGCGCTGCTGCGCTTATATGCGCAATCGCGATC 348
Oy 155 GlyGlyLeuIleAanIleValThrLySerAseLeuGluGluGlnPheGluThrArg 174
Db 347 GCGGCGGTGTCATCATCATCATCATCCGCTTGACGCGCGCGCTCGGCGCAATTCACG 288
Oy 175 IleGlyValHisGlySerLySerLySerSerGluGlyIleGlyTyGlnAlaGlyGlnSer 194
Db 287 GTGGGC-----GGCGATCGTTCCAGACGATCGCTTCGGGCGTACGCTGCGCGCAT 234
Oy 195 ValAlaGly-----ValSerGluAanGlyAanVal 204
Db 233 CTGGGGGCGCGCGACACTTCGACCTCGACCTCGGCTATGTCGCGCAGAACGGC----- 180
Oy 205 LeuAlaArgLeuAseValAseTyArgThrThrGlyGlyAlaPheAseAlaAanGlyLyS 224
Db 179 -----GAGTTCCGACCGCGCGACGGCGCTGATCCGACCGAAACAGCGAC 138
Oy 225 ArgIleAlaProGluProAlaGlnThrAseLyGlnAseSer 238
Db 137 TCGATGCGGGCGTCCGG-CCGGGAGACGATCGCGCGGACGCT 97

RESULT 7
LOCUS AZ302660/c 519 bp DNA linear GSS 06-MAR-2001
DEFINITION GSSBrut1597 Brucella abortus random genomic library Brucella
melitensis biovar Abortus genomic clone U01597, DNA sequence.
ACCESSION AZ302660
VERSION AZ302660.1 GI:10128871
KEYWORDS GSS.
SOURCE Brucella melitensis biovar Abortus.
ORGANISM Brucella melitensis biovar Abortus
Brucella melitensis biovar Abortus
Brucellaceae; Brucella.
1 (bases 1 to 519)
REFERENCE Sanchez,D.O., Zandemont,R.O., Cravero,S., Verdun,R.E., Pierron,E.,
Paccio,P., Diaz,G., LanzaVecchia,S., Agniero,F., Fraech,A.C.C.,
Anderson,S.G.E., Rosetti,O.L., Grau,O. and Ugalde,R.A.
Gene discovery through genomic sequencing of Brucella abortus
infect. Immun. 69 (2), 865-868 (2001)
21101034
TITLE Contact: Siv Andersson
JOURNAL Small Genomes Sequencing Group
MEDLINE Department of Molecular Evolution, Uppsala University
COMMENT Norrbygsen 18C, S-752 36, Uppsala, Sweden
Tel: 46-18-471-4379
to biotinylated driver (prepared from the same library by
PCR) to Cot-omega of 11. After removal of hybrids and
excess driver by streptavidin sepharose chromatography,
the ss-phagemids were made double stranded and
electroporated into Top-10 F'. Original library
construction by Bruce Blumberg (Cho et al. 1991 Cell 67,
1111-1120). Note: This is a Xenopus Gene Collection (XGC)
library."

```

Fax: 46-18-471-6404  
Email: Siv.Anderson@ebc.uu.se  
Sequences were basecalled with phred and vector was masked with  
crossmatch (see http://genome.washington.edu). Sequences were then  
trimmed from both ends to remove low quality bases and masked  
vector.  
Class: shotgun.

FEATURES  
source  
Location/Qualifiers  
1..519  
/organism="Brucella melitensis biovar Abortus"  
/strain="2308"  
/db\_xref="taxon:235"  
/clone="UUI597"  
/clone\_lib="Brucella abortus random genomic library"  
BASE COUNT 111 a 164 c 140 g 104 t  
ORIGIN  
/note="Vector: modified M13"

Alignment Scores:  
Pred. No.: 0.00138 Length: 519  
Score: 120.00 Matches: 50  
Percent Similarity: 36.98% Conservativeness: 21  
Best Local Similarity: 26.04% Mismatches: 61  
Query Match: 5.86% Indels: 60  
Gaps: 8

US-09-889-746-2 (1-400) x AZ302660 (1-519)

Qy 91 GlyValSerSerGlyThrThrSerAsnPheGlyGlnThr----- 103  
Db 498 GGCATTTCGTCACAGACGATCGCGGCGAGGGTCTTCGCCGATATCTATATGCGCGC 439  
Qy 104 MethHisGlyArgGlnValGlnPheLeuLeuAanGlyValProLeuThrGlySerArgAsp 123  
Db 438 ATGTCGTCGAAGCAGACGCTGCTGCTGCTCAATGGGTCGCGACGCTTCGCGAACGACT 379  
Qy 124 IleSerArgGlnLeuAanSerIleAanProAanGlnValAlaAargIleGluValLeuSer 143  
Db 378 GGATCGACGCGCTTCGCCAATATTCGCTGACCTCCATCGAAGCGTATCGAGATTGCCAGG 319  
Qy 144 GlyAla---ThrSerIleTyGlySerGlyAlaThrGlyGlyLeuIleAanIleValThr 162  
Db 318 GCGCGCATCTCCGCAATATGCGCGGATGCATCGCGGGTGTCAATATCATCATCACC 259  
Qy 163 LysSerAspLeuGluGluGlnPheGluThrArgIleGlyValHisGlySerLysLeu 182  
Db 258 AACACGGGC-----GGGCGCTGCGCGCAACGCGCA 229  
Qy 182 ----- 182  
Db 228 TGGTGGCGAGCGTTTCAACAGCGGTGTCGATCCATGGGCGGTATGATCGGGTTCCG 169  
Qy 183 -----SerSerGlyIleGlyTyGlnValGlyGlnSerValAlaGlyVal 198  
Db 168 TTACAAGCGCGCAGCAGCGATATGATGATGATGATGATGATGATGATGATGATGATGAT 109  
Qy 199 SerGluAanGlyAanValLeuAlaArgLeuAanValAspTyArgThrThrGlyGlyAla 218  
Db 108 CAA-----GGCTATGATTTC---ACCACGCGCGAAGCA 79  
Qy 219 PheAspAlaAanGlyLysArgIleAlaProGluProAlaGlnThrAsp---LysGlnAsp 237  
Db 78 TTC-----GCCACGAGCGCGGACGATGATGATGATGATGATGATGATGATGATGATGAT 40  
Qy 238 SerLysSerLeuSerValAanThrAanValAspTyr 249  
Db 39 TCCTTCAATTTACGCTGTGCGAAGATTTCGACTCG 4

RESULT 8  
BH397318 538 bp DNA linear GSS 11-DEC-2001  
LOCUS AG-ND-135B5.TR ND-TAM Anopheles gambiae genomic clone AG-ND-135B5,  
DEFINITION DNA sequence.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

BH397318  
BH397318.1 GI:17343534  
GSS.  
African malaria mosquito.  
Anopheles gambiae  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Nematocera; Culicoides;  
Anopheles.

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 538)  
Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.  
Direct Submission of BAC-end sequences from Anopheles gambiae  
Unpublished (2001)  
Other\_GSSs: AG-ND-135B5.TP  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0208  
Fax: 301 838 3543  
Email: bjoftus@tigr.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by  
F.H. Collins and sequenced by The Institute for Genomic Research  
(TIGR). The BAC library was generated from A. gambiae PEST strain  
DNA. All DNA was extracted from newly hatched first instar larvae  
to minimize the inclusion of DNA from microorganisms that inhabit  
the gut. The DNA is derived from mixed sexes of larvae. The BAC  
library was constructed at Texas A&M University BAC Center  
University, College Station, Texas 77843-2123, USA using a HindIII  
partial digest.  
Seq primer: M13 Rev  
Class: BAC ends.

FEATURES  
source

Location/Qualifiers  
1..538  
/organism="Anopheles gambiae"  
/strain="PEST"  
/db\_xref="taxon:7165"  
/clone="AG-ND-135B5"  
/clone\_lib="ND-TAM"  
/note="vector: pECBAC1; Site\_1: HindIII"

BASE COUNT 178 a 98 c 122 g 140 t

Alignment Scores:

Pred. No.: 0.00741 Length: 538  
Score: 114.00 Matches: 48  
Percent Similarity: 40.96% Conservativeness: 29  
Best Local Similarity: 25.53% Mismatches: 61  
Query Match: 5.57% Indels: 50  
Gaps: 9

US-09-889-746-2 (1-400) x BH397318 (1-538)

Qy 71 GluGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeu 90  
Db 24 CAGCGTGTAGTGCATCAGCACACCGCGTGAACAGCGCGGCGTAAGATTAGAGAAATT 83  
Qy 91 GlyValSerSerGlyThrThrSerAsnPheGlyGlnThrMetHisGlyArgGlnValGln 110  
Db 84 GGCCTATAAGCGCATCCAGTTCCTCGCTT----- 113  
Qy 111 PheLeuLeuAanGlyValProLeuThrGlySerArgAspIleSerArgGlnLeuAanSer 130  
Db 114 TATGTAGTAGCGGTATTCCTTACGATCGAAT-----ATAATGCT 155  
Qy 131 IleAanProAanGlnValAlaAargIleGluValLeuSerGlyAlaThr-----SerIle 148  
Db 156 ATAAATCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 215  
Qy 149 TyrGlySerGlyAlaThrGlyGlyLeuIleAanIleValThrLysSerAspLeuGlu 168  
Db 216 TATGTTCCAGAGGTGCAATGGAATTATTGTAATTACGACTAAATCCGGAAGAAAGGA 275  
Qy 169 Glu-----GlnPheGluThrArgIleGlyValHisGlySer----- 180

Db 276 GAGCTAAATTAACCTAATTAAGCCAGGATTTTCCGGAGAGCTGTAAAGATTAT 335  
 Qy 181 -----LysLeuSerSerGluGlyIle-----GlyTyr 189  
 Db 336 AAACAGTAATATACAGATCATTTTCAGTTATACGTGGAAAGCTTAAAGAAAGCTTAT 395  
 Qy 190 -----GlnValGlyGlnSerValAlaGlyValSerGluAsnGlyAsnValLeuAla 206  
 Db 396 GTATCTCTCTAAATTTCTCAGAGCCAGCTGCACAAATGCAACAGACATCTTAATTTCC 455  
 Qy 207 ArgLeuAspValAspTyrArgThrThrGlyAlaPheAspAlaAsnGlyValArgIle 226  
 Db 456 AGTTAGGAATTAAT-----CCATACGCTTCAGCTTAT----- 488  
 Qy 227 AAlaProGluProAlaGlnThrAsp 234  
 Db 489 ---CCTAAACCGCTTGCTGTGAT 509

RESULT 9  
 BH382603 558 bp DNA linear GSS 10-DEC-2001  
 LOCUS AG-ND-102C22.TR ND-TAM Anopheles gambiae genomic clone AG-ND-102C22  
 DEFINITION  
 ' DNA sequence.  
 ACCESSION BH382603.1 GI:17328745  
 VERSION BH382603.1  
 KEYWORDS GSS.  
 SOURCE African malaria mosquito.  
 ORGANISM Anopheles gambiae  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
 Anopheles.  
 1 (bases 1 to 558)  
 Shetty,J.J., Malek,J.J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.  
 TITLE Direct Submission of BAC-end sequences from Anopheles gambiae  
 JOURNAL Unpublished (2001)  
 COMMENT Other\_GSSs: AG-ND-102C22.TF  
 Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: b.loftus@igf.org

REFERENCE  
 AUTHORS Shetty,J.J., Malek,J.J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.  
 TITLE Direct Submission of BAC-end sequences from Anopheles gambiae  
 JOURNAL Unpublished (2001)  
 COMMENT Other\_GSSs: AG-ND-102C22.TF  
 Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: b.loftus@igf.org  
 This clone is from an A. gambiae BAC library (ND-TAM) provided by F.H. Collins and sequenced by The Institute for Genomic Research (TIGR). The BAC library was generated from A. gambiae PBST strain DNA. All DNA was extracted from newly hatched first instar larvae to minimize the inclusion of DNA from microorganisms that inhabit the gut. The DNA is derived from mixed sexes of larvae. The BAC library was constructed at Texas A&M University BAC Center University, College Station, Texas 77843-2123, USA using a HindIII partial digest.  
 Seq primer: M13 Rev  
 Class: BAC ends.

FEATURES  
 source Location/Qualifiers  
 1..558  
 /organism="Anopheles gambiae"  
 /strain="PBST"  
 /db\_xref="taxon:7165"  
 /clone="AG-ND-102C22"  
 /clone\_1lb="ND-TAM"  
 /note="Vector: pECBAC1; Site\_1: HindIII"

BASE COUNT 179 a 96 c 130 g 153 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.0103 Length: 558  
 Score: 113.00 Matches: 48  
 Percent Similarity: 40.96% Conservatave: 29  
 Best Local Similarity: 25.53% Mismatches: 61  
 Query Match: 5.52% Indels: 50  
 DB: 17 Gaps: 9

US-09-889-746-2 (1-400) x BH382603 (1-558)  
 Qy 71 GlnGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeu 90  
 Db 24 CAGCGTGAATGTCATCGACGACACCCCGGTAAACAGCGGCGATTAAGAAATT 83  
 Qy 91 GlyValSerSerGlyThrThrSerAsnPheGlyGlnThrMetHisGlyValGln 110  
 Db 84 GGCTCTAATAGCGCATTCACAGTTCTCCGCT----- 113  
 Qy 111 PheLeuLeuAsnGlyValProLeuThrGlySerArgAspIleSerArgGlnLeuAsnSer 130  
 Db 114 TATGTAGTAGACGGATTAATCTTACAGATGAAT-----ATAAATGCT 155  
 Qy 131 IleAsnProAsnGlnValAlaArgIleGluValLeuSerGlyAlaThr-----SerIle 148  
 Db 156 ATAAATCTTAATGATGTACTTATTAACGTGTCTTAAAGATGCAACGCAAGTCACTG 215  
 Qy 149 TyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrIleSerAspLeuGlu 168  
 Db 216 TATGTTCCAGAGCGTCAAAATGCAATTTGTAATTAAGCACTAAATCCGGAAGAAAGGA 275  
 Qy 169 Glu-----GlnPheGlnThrArgIleGlyValHisGlySer----- 180  
 Db 276 GAGCTAAATTAATTAACCTCAATTAAGCCAGGATTTTCCGGAGAGCTGTAAAGATTAT 335  
 Qy 181 -----LysLeuSerSerGluGlyIle-----GlyTyr 189  
 Db 336 AAACAGTAATATACAGATCATTTTCAGTTATACGTGGAAAGCTTAAAGAAAGCTTAT 395  
 Qy 190 -----GlnValGlyGlnSerValAlaGlyValSerGluAsnGlyAsnValLeuAla 206  
 Db 396 GTATCTCTCTAAATTTCTCAGAGCCAGCTGCACAAATGCAACAGACATCTTAATTTCC 455  
 Qy 207 ArgLeuAspValAspTyrArgThrThrGlyAlaPheAspAlaAsnGlyValArgIle 226  
 Db 456 AGTTAGGAATTAAT-----CCATACGCTTCAGCTTAT----- 488  
 Qy 227 AAlaProGluProAlaGlnThrAsp 234  
 Db 489 ---CCTAAACCGCTTGCTGTGAT 509

RESULT 10  
 BH001150 428 bp DNA linear GSS 01-JUL-2001  
 LOCUS BH001150  
 DEFINITION A2 Pirellula marina lambda zap Express Library Pirellula marina genomic clone A2 similar to vitamin B12 receptor precursor, DNA sequence.  
 ACCESSION BH001150  
 VERSION BH001150.1 GI:14579916  
 KEYWORDS GSS.  
 SOURCE Pirellula marina.  
 ORGANISM Pirellula marina  
 Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;  
 Planctomycetaceae; Pirellula.

REFERENCE  
 1 (bases 1 to 428)  
 Jenkins,C., Kedar,V. and Fuerst,J.A.  
 TITLE Gene discovery from sequence tags generated using genomic DNA libraries constructed from representatives of the planctomycete division of the Domain Bacteria  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Fuerst JA  
 Department of Microbiology and Parasitology  
 University of Queensland  
 Brisbane, QLD 4072, Australia  
 Tel: +617 3365 4643  
 Fax: +617 3365 4620  
 Email: fuerst@biosci.uq.edu.au  
 Class: Shotgun.  
 Location/Qualifiers  
 1..428  
 /organism="Pirellula marina"

FEATURES  
 source

```

/strain="ACM 3344"
/db_xref="taxon:124"
/clone="A2"
/clone_lib="Pirellula marina Lambda Zap Express Library"
/notes="Vector: Lambda Zap Express BamHI arms; Site 1:
BamHI; Site 2: BamHI; Purified genomic DNA from Pirellula
marina was restricted with BclI to give fragments of 2-9kb
and ligated into the BamHI sites of the Lambda Zap Express
vector. The ligated DNA was packaged into Gigapack Gold
III phage heads. E.coli XL1-Blue MRP' was infected with
the packaged phage. Recombinant phage plaques were picked
and stored in phage storage buffer (PSB). Phage/PSB was
used as template for PCR."
BASE COUNT 101 a 119 c 131 g 77 t
ORIGIN

Alignment Scores:
Pred. No.: 0.00762 Length: 428
Score: 112.50 Matches: 32
Percent Similarity: 50.94% Conservatives: 22
Best Local Similarity: 30.19% Mismatches: 47
Query Match: 5.50% Indels: 5
DB: 17 Gaps: 3

US-09-889-746-2 (1-400) x BH001150 (1-428)
Qy 67 GluGlnIleGln-----GluGlnAlaThrGlySerArgGlnLeuAlaAapValMetAla 84
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 64 GAAGAGATTCAACGGACCCCAACAGACGACCGTGGCCGAGTAGTCTGGCGAGGTCGAAGGG 123
Qy 85 GlnLeuIleProSerLeuGlyValSerGlyThrThrSerAsnPhelGlnThrMet 104
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 124 GTTGATGTGTTTCGACCGCGCGCGCGCGCGGATCAGCTCGTCTTTATCCGCGCGGCC 183
Qy 105 HisGlyArgGlnValGlnPheLeuLeuAanglyValProLeuThrGlySerArgAepIle 124
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 184 AATTCCGACATACCAAGGTAATTCGACGGCATTCGATCCATCAACGATCCCGACAGGCC 243
Qy 125 SerArg-----GlnLeuAanSerIleAanProAanGlnValAlaAargIleGluValLeu 142
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 244 GGACGAAGCTTTGACTCTCGAGTGTTCGACGCAAAACATCGAGCGGATCGAAGTGTG 303
Qy 143 SerGlyAlaThrSer---IleThrGlySerGlyAlaThrGlyGlyLeuIleAanIleVal 161
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 304 CGCGGCCCGCAAGACATGTTGTATGCTCGAGCGCATCGCGCGCGTGTATCCAAAGAAAT 363
Qy 162 ThrLysSerAspLeuGlu 167
|||:|||||:
Db 364 CAAAAGCTTCTCGAGAG 381

RESULT 11
P132R/c
LOCUS
DEFINITION
Leishmania major Friedlin PAC P132 right end-sequence, similar to
AF051693 AF051693 Pseudomonas aeruginosa... N=234, Prob=1.6e-24;
TR:068590 068590 HYDROXAMATE-TYPE FERRISIDEROPHORE... N=150,
Prob=6.1e-30, genomic survey sequence.
ACCESSION
AL160522
VERSION
AL160522.1 GI:7258925
KEYWORDS
GSS.
SOURCE
Leishmania major.
Leishmania major
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE
1 (bases 1 to 423)
Ivens,A.C., Lewis,S.M., Bagherzadeh,A., Zhang,L., Chan,H.M. and
Smith,D.C.
TITLE
A physical map of the Leishmania major Friedlin genome
JOURNAL
Genome Res. 8 (2), 135-145 (1998)
MEDLINE
98146435
PUBMED
9477341
REFERENCE
2 (bases 1 to 29)
Taylor,R.G., Huckle,E.E.J., Ivens,A.C., Rajandream,M.A. and
AUTHORS
Barrell,B.G.
Direct Submission
Submitted (14-MAR-2000) Leishmania major Friedlin genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
alicat@sanger.ac.uk
see http://www.ebi.ac.uk/parasites/leish.html
Details of Leishmania sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/L_major/
The primer sequence can be obtained from alicat@sanger.ac.uk.
FEATURES
Location/Qualifiers
source
1..423
/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/clone="PAC P132"
BASE COUNT 63 a 155 c 135 g 70 t
ORIGIN

Alignment Scores:
Pred. No.: 0.00856 Length: 423
Score: 112.00 Matches: 41
Percent Similarity: 42.38% Conservatives: 23
Best Local Similarity: 27.15% Mismatches: 65
Query Match: 5.47% Indels: 22
DB: 17 Gaps: 7

US-09-889-746-2 (1-400) x P132R (1-423)
Qy 54 IleThrGlnMetProHisThrThrLysValIleTyThrGluGlnIleGlnGluAla 73
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 415 CTGCGCGAGTGGCCGACACGCTCAACGTGTGTGACGGCGGAGTCTATCGGCACGACAC 356
Qy 74 ThrGlySerArgGlnLeuAlaAapValMetAlaGlnLeuIleProSerLeuGlyValSer 93
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 355 GCGGCTCGATCCAG-----GACACGCTGAAGAAAT---GTGCCGGGGTGGCTTCG 305
Qy 94 SerGlyThrThrSerAsnPhelGlnThrMetHisGlyArgGlnValGlnPheLeuLeu 113
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 304 CATGGG-----GATGACACGCGCGACGACGATGATGATC 272
Qy 114 AanglyValPro-----LeuThrGlySerArgAspIleSerArgGln 127
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 271 CGCGGCTTCCCGCATCGCCGACGATCTGTCGACGCGATCCGGGACGCGCTGTAT 212
Qy 128 LeuAanSerIleAanProAanGlnValAlaAargIleGluValLeuSerGly---AlaThr 146
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 211 TTCCGCGATCTG-----TCGAACGTCGAGCGGATCGAGGTGATCAAGGACCGCGCC 158
Qy 147 SerIleTyThrGlySerGlyAlaThrGlyLeuIleAanIleValThrLysSer---Asp 165
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 157 GTGCTGTACGCGCGCGGTCTGTCGGCGGCTGATCAACCGCGTGCACCAAGACGCGGC 98
Qy 166 LeuGluGluGluGlnPheGlnThrArgIleGlyValHisGlySerLysLeuSerGlu 185
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 97 CTGGACATCACCAGTCTGAGCTCAGTTCAGGCTGTGGGCGGACGCGCGCGAGGTC 38
Qy 186 GlyIleGlyTyThrGlnValGlyGlnSerValAla 196
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 37 GACGTCCGCGCGAGTATTTCGGCGATGCGCGCGCC 5

RESULT 12
BH825061
LOCUS
DEFINITION
BACPP21-O12.Y Pristionchus pacificus BAC ends Pristionchus
pacificus genomic, DNA sequence.
ACCESSION
BH825061
VERSION
BH825061.1 GI:21004090
KEYWORDS
GSS.
SOURCE
Pristionchus pacificus.
Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.

```





```

gene
misc_feature
1..3064
/gene="Sp3"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev stage="adult"
1..3064
/gene="Sp3"
1..3064
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev stage="adult"

BASE COUNT 977 a 577 c 653 g 857 t
ORIGIN

Alignment Scores:
Pred. No.: 1.49 Length: 3064
Score: 105.00 Matches: 87
Percent Similarity: 34.01% Conservative: 63
Best Local Similarity: 19.73% Mismatches: 173
Query Match: 5.13% Indels: 118
DB: 11 Gaps: 18

US-09-889-746-2 (1-400) x AK004607 (1-3064)
QY 18 A1aValThrGlnGlnLeuTyraGlnProAsnGluSerLeuProThrValGlnGlnGln 37
DB 243 GGTACTTCAGAGGGGAGATGTCCTCCCTTCAGAAATTCAGATTCAGATTCAGAAATATTT 302
QY 38 ProValValIleThrIleAspLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMet 57
DB 303 TCAGTTGCACAGATCATCATTCATCAATGCGACAGTGTCCAT----- 347
QY 58 ProHisThrThyLeuValIleTyrglnGlnGlnIleGlnGlnGlnAlaThrGlySerArg 77
DB 348 -----GTCAATGATTCAGATTCATTCACCAATTCAGTCAACAGAGCTCAG 392
QY 78 GlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSerSerGlyThrThr 97
DB 393 CAGGTT-----CAGATTGGCTTC--ACAGCTCTCTCA 422
QY 98 SerAsnProGlyGlnThrMetHisGlyArgGlnValGln----- 110
DB 423 GATTAATGGGGCATTAATCAAGAAACAGCCAAATTCAGATTCATCTCGGCTTAATCAA 482
QY 111 ----PheLeuLeuAsnGlyValProLeuThrGlySerArgAspIleSerArgGlnLeuAsn 129
DB 483 ACCTTACTGCGCTCTGAAACCTCTGCTAATATCCAGATTCATACCAAGACTGGT 542
QY 130 SerIleAsnProAsnGlnValAlaArgIleGlnValLeuSerGlyAlaThrSerIleTy 149
DB 543 CAAAGTCAGGTTCAAGGAGTTGCA-----ATT 569
QY 150 GlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLysSerAspLeuGlnGln 169
DB 570 GGTTGGCTCATATTCCTGCTCAACTCAAGTAGTGCCTAATGCTGCTGCTGCCA 629
QY 170 GlnPheGlnThrArgIleGlyValHisGlySerLysLeuSerSerGlyIleGlyTy 189
DB 630 GGAATATTAATACCTTTGACCAATCAATAGTGCATCTGATCTTTGGGACTCTGGGA 689
QY 190 GlnValGlyLysSerValAlaGlyValSerGluAsnGlyValAsnValLeuAlaArgLeuAsp 209
DB 690 AGTTTCACAGACATGACTGCAGGACATTAATGCCGATGGACATTGGATA----- 737
QY 210 ValAspTyrArgThrThrGlyGlyAlaPheAspAlaAsn-----Gly 223
DB 738 -----AACACAGGACAGCTATGATGATGATTCAGACAAATTCAGAAAGACTGGT 785
QY 224 LysArgGlyIleAlaProGluProAlaGlnThrAspLysGln-----Asp 237
DB 786 GAGCGGGTTCTCTCATGATGTTAATGAACATTAATGCTGATACAGATTATTATTGTCACA 845

```

```

QY 228 SerLysSerLeuSerValAsnThrAsnValAspTyrGlnLeuAspLysGlnAsnIle 257
DB 846 TCCTCTTCATCAGACTGCTGCTTACATATGATAGTATATTCAGCAAAACACA 905
QY 258 AsnLeuAlaLeuThr-----HisTyrAsnAspLysGlnAspThrAspTyr 272
DB 906 AATTAAGCTGACTACTACAGTGGGCAAGTCCATTTCTCAGATCTTCAG----- 953
QY 273 AlaProAspTyrGlyValAsnArgLeuAlaValLeuPheGlyGlyLysProSerLeuAsnAla 292
DB 954 -----GAAATTAATATTCAGTCGCTGCTTTCTGAAAGACACAGCTCAGAT 1001
QY 293 IleLys-----GlyLeuSerLeuSerGlnGln 301
DB 1002 ATTCAGATTTCACAGACAGAGCTGTGTACAAATTCATCAAGATTCACAGAG 1061
QY 302 ProLysThr-----ThrLysSerThrPheAsnIleAsnTyr 313
DB 1063 CCACACAGTCAAGCCCAATGTCGAGATTAATACACACAGCAATC----- 1109
QY 314 HisHisAspAspLeuThrTyrGlyValAsnThrIleAsn-----ThrAsnAlaTyrTyr 329
DB 1110 CATGGCGTGCAGAGCCAGTGTCAAAATATATATACACACAGGCTTCGAAACCTTCAGTTG 1169
QY 330 ArgArgGlyLysGlyArgPheTyr-----ProPheValAlaPro----- 342
DB 1170 CAGCTGAATTCCTGGAACCTTTTATATTCAGCAGACAGACAGTACCCCTTCGACAGATA 1229
QY 343 -----PheSerIleAlaLysAlaLeuProIleLeuGlnSerMetAsnLeuPro 358
DB 1230 ACTTGGCAACATTCACAGTA---CAAGGGGTCCAGAACTTACAGAAATTCGAAATACAA 1286
QY 359 SerAlaThrLeuAspAlaTyrThrLysAlaPro-----GlnAla 371
DB 1287 AATTAAGCTGCCCCAACAATTAATTTAGAGCTGTTCAGACACTACAGCTGCTCAAGTC 1346
QY 372 ArgAlaTyrGlyValLeuGlnSerLysSerLysAlaGlnValLeuGlyArgValProAsn 391
DB 1347 GCAGCAGGTGGAGCCTTGACTTCACTCACTCAAGTCAAGTCAAGTCAAGTTCGCAAT 1406
QY 392 Leu 392
DB 1407 CTA 1409

RESULT 15
AG989725/c 694 bp DNA linear GSS 14-AUG-2000
LOCUS Rf00358 Photorhabdus luminescens strain W14 M13 library
DEFINITION Photorhabdus luminescens genomic clone PLG00358, DNA sequence.
ACCESSION AG989725
VERSION AG989725.1 GI:9648319
KEYWORDS GSS.
SOURCE Photorhabdus luminescens.
ORGANISM Photorhabdus luminescens.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Photorhabdus.
REFERENCE 1 (bases 1 to 694)
AUTHORS French-Constant,R.H., Waterfield,N., Burland,V., Perna,N.T.,
Daborn,P.J., Bowen,D. and Blattner,F.R.
TITLE A genomic sample sequence of the entomopathogenic bacterium
Photorhabdus luminescens W14: potential implications for virulence
JOURNAL Appl. Environ. Microbiol. 66 (8), 3310-3329 (2000)
MEDLINE 20378633
COMMENT Contact: french-constant RH
Department of Biology and Biochemistry
University of Bath
South Building, Bath BA2 7AY, UK
Tel: (44) 1225 826621
Fax: (44) 1225 826779
Email: bsr10@bath.ac.uk
This is one of 2,122 random reads from the M13 library. For
annotation of identified clones (BLASTX, BLASTN and mapping to E.
coli K12 genome) please see french-constant et al. 2000, Nucleic

```

```
Acids Res.
Seq primer: M13 Forward
Class: shotgun.
FEATURES
    source
        1..694
            /organism="Photorhabdus luminescens"
            /strain="W14"
            /db_xref="taxon:29488"
            /clone="PLG00358"
            /clone_lib="Photorhabdus luminescens strain W14 M13 library"
            /dev_stage="primary phase variant"
            /note="genomic DNA from strain W14 was size selected (1-2 kb) and then cloned into M13 Janus."
BASE COUNT    204 a 159 c 128 g 195 t      8 others
ORIGIN
Alignment Scores:
Pred. No.:      0.147      Length:      694
Score:          104.50     Matches:     46
Percent Similarity: 41.07%   Conservative: 23
Best Local Similarity: 27.38% Mismatches:    60
Query Match:      5.11%     Indels:      39
Db:              17        Gaps:       7
US-09-889-746-2 (1-400) x AQ989725 (1-694)
QY 120 GlySerArgAspIleSerArgGlnLeuAenSerIleAenProAenGlnValalaAargIle 139
Db 539 GGCTCTTTGATTTAAGCCAG-----ATACCTGTATCATTTGTTTCAGAAATC 492
QY 140 GluValLeuSerGlyAlaThrSer---IleTyrGlySerGlyAlaThrGlyGlyLeuIle 158
Db 491 GAATATATCCGGCGACCATCTCCGCTGTTTATGTTCTGATGCTATTGCTGTTGTTATT 432
QY 159 AsnIleValThrLysSerAspLeuGluGluGlnPheGluThrArgIleGlyValHis 178
Db 431 AATATCATCATCAACAGAGAAACTAGGCGCTAACCTTAACGTAGGTATCGTTCCAT 372
QY 179 GlySerLysLeuSerSerGluGlyIleGlyTyrGlnVal-----GlyGlnSerValala 196
Db 371 -----GGTTATCAGACTTATGATGCGCGACTCAGCAA 339
QY 197 GlyValSerGluAenGlyAenValLeuAlaAargLeuAsp----- 209
Db 338 AAGTTAGCTGAAATAACAGCTCCTGACAGCAGCAGCTAACTATACCTATATAAGGATAC 279
QY 210 -----ValAspTyrArgThrThrGlyAlaPheAspAlaAenGlyLysAargIleAla 227
Db 278 GATGTTGTTGCTTATGTTATGTAATCTGTTGCTTT----- 246
QY 228 ProGluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAenThrAsnVal 247
Db 245 -----AGACAGCTGATCGTACGCGTTTATAGAGCAAAATGCTTGGTTGGGAGTC 195
QY 248 AspTyrGlnLeuAspAspLysGlnAenIleAenLeuAlaLeuThrHisTyrAsnAspLys 267
Db 194 GATCAAAATTTAATGAGCAATACAGTGGGTTTATTCGAGCTTATGCTATACAATAAA 135
QY 268 GlnAspThrAspTyrAlaProAsp 275
Db 134 -----ACATCTTATGTCGGAT 117
RESULT 16
B1158313
LOCUS      B1158313      839 bp      mRNA      linear      EST 05-JUL-2001
DEFINITION 602921002F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5061072 5',
            mRNA sequence.
ACCESSION  B1158313
VERSION    B1158313.1 GI:14618134
KEYWORDS   EST.
SOURCE     house mouse.
ORGANISM   Mus musculus
```

```
REFERENCE
AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Straubeberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
            DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM1165 row: d column: 01
            High quality sequence stop: 782.
FEATURES
    source
        1..839
            /organism="Mus musculus"
            /strain="129,C57BL/6J,FVB/N"
            /db_xref="taxon:10090"
            /clone="IMAGE:5061072"
            /clone_lib="NCI_CGAP_Mam3"
            /tissue_type="tumor, gross tissue"
            /dev_stage="10 months"
            /lab_host="DH10B"
            /note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI;
            Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt.
            Library constructed by Life Technologies Investigators.
            providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
            Reference for transgenic model: Xu et al., Nature Genetics
            22, 37-43 (1999)."
```

BASE COUNT 232 a 252 c 208 g 147 t

ORIGIN

Alignment Scores:  
Pred. No.: 0.395 Length: 839  
Score: 102.00 Matches: 35  
Percent Similarity: 45.86% Conservative: 37  
Best Local Similarity: 22.29% Mismatches: 54  
Query Match: 4.98% Indels: 31  
Db: 13 Gaps: 6

US-09-889-746-2 (1-400) x B1158313 (1-839)

QY 5 HisTyrPhe-GlnTyrLeuSerLeuProLeuLeuSerValAlaValThrGlnGlnLeuTy 24  
Db 136 TACTACTTCACCTGGCTCAGCAGCCACTGCTCCAGGTCCAGGACAGGACGACGAGGC 195  
QY 24 zAlaGlnProAenGluSerLeuProThrValGluLeuGluProValVal-----Il 41  
Db 196 CAGCAGACCA-----CCAGTTCTACGACCACCATC 225  
QY 41 eThrIleAspLysSerGlyMetAlaLeuAlaAenArgIleThrGlnMetPro----- 58  
Db 226 CAGCCTGGACAGATCATTCATTCAGCAGCCTCAGCAGGCTCAGACACACCCCGTGC 285  
QY 59 -----HisThrThrIleValIleTyrGluGlu-----GlnIleG 70  
Db 286 CAGTTGGAGGAGGTGACAGGTGCGATTGTGACGGCCCACTCAGGGTTCAGGCCCA 345  
QY 70 nGluGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLe 90  
Db 346 GCAGACCCAGAGTGGCACTGGACAGACCATGCAGGTGATGCAGCAGATCATTT----- 397  
QY 90 uGlyValSerSerGlyThrThrSerAenPheGlyGlnThrMetHisGlyArgGlnValG 110  
Db 398 ----ACCAACACAGGAGAGATCCAGCAGATCCCGGTGCGAGCTGAATGCTGGCCAG 453  
QY 110 nPheLeuLeuAenGlyValProLeuThrGlySerArgAspIleSerArgGlnLeuAenSe 130  
Db 454 ATATATCCGATTAGCCCGCTGTATCAGGACCCCAAGTTGTTCAGGACAGATCCAGAC 513



QY 130 r1e-----AenProaenglnvala1aargillegluval1leuSergly 144  
 Db 514 ACTTGCTACCAATGCCAACAGATCACACAGACAGAGCTCCAGCAAGA 562

RESULT 17  
 BC017847

LOCUS BC017847 1344 bp mRNA linear HTC 06-DEC-2001  
 DEFINITION Homo sapiens, clone IMAGE:4288585, mRNA.  
 ACCESSION BC017847  
 VERSION BC017847.1 GI:17389644  
 KEYWORDS HTC.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 1344)  
 REFERENCE Direct Submission  
 TITLE Strausberg, R.  
 AUTHORS Submitted (03-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA  
 JOURNAL NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 CONTACT: MGC help desk  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: CLONTECH  
 CDNA Library Preparation: CLONTECH Laboratories, Inc.  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
 DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305  
 Web site: <http://www-shgc.stanford.edu>  
 Contact: (Dickson, Mark) [mcddpaxil.stanford.edu](mailto:mcddpaxil.stanford.edu)  
 Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>  
 Series: ILNL Plate: 36 Row: a Column: 3  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7662155  
 This clone has the following problem: frame shifted.

FEATURES  
 source  
 1..1344  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4288585"  
 /cissue\_type="Skeletal Muscle"  
 /clone\_id="NIH MGC\_81"  
 /lab\_host="DH10B"  
 /note="Vector: pDNR-LIB"

BASE COUNT 500 a 273 c 278 g 293 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 0.86 Length: 1344  
 Score: 102.00 Matches: 100  
 Percent Similarity: 35.62% Conservative: 71  
 Best Local Similarity: 20.83% Mismatches: 172  
 Query Match: 4.98% Indels: 18  
 Db: 11 Gaps: 20

US-09-889-746-2 (1-400) x BC017847 (1-1344)

QY 16 SerValAlaValThrcInglInleuTYrAlaGlnProaenglnSerleupProthrValGlu 35  
 Db 13 TCACTGGGCTCTCTCAAGTGTCTGTTTCCCACTTCATCCCAATTCACGCTGTATAT 72

QY 36 LeugluProValValIleThrIleAspIlySergly-MetAlaLeuAlaAsnArg----- 53  
 Db 73 CTACAAACAGTTGCTGATGAGTTTCTGACAAAGCGGAGCAAAATTAATGACACGGCGACA 132

QY 54 -----1leth 55

Db 133 ACATGAGAGAGCTCTAGAGTGAAGCAGCTTAGTGCAGAGAAGACAGAAATTG 192  
 QY 55 rglInet-----ProthIlethr 61  
 Db 193 TCAATTCGAAAAACAACCTTTGGCTGCTGGGACAAAGATTAATTAACCCAAAATCC 252

QY 61 rlyValIleTyrgluGlnIleGlnIleuAlaThrIlySerglyArgGlnleuAlaAs 81  
 Db 253 TAAG-----AAAGAACTGGAGGACAGAGAACAGAAACAGAAATATGCAAG 300

QY 81 pValMetAlaGlnleuIlePro----- 88  
 Db 301 TGAAGAGAAATCTCCAGTACCTCTGACTCATCTTAAACAGTGAAGCTCCATTCACA 360

QY 89 -----SerleuGlyValSerglyThrIleSerleuPheGlyThrMetI 105  
 Db 361 AGAATTAGGACCCCTGCTGTGAAATATGTACCATCCGAGCTTAATGACAGAGACGCC 420

QY 105 eglYArgGlnValGlnPheleuAsnGlyValProleuThrIlySerglySergAspIle 125  
 Db 421 AGGAGTCCAGATCAAGATAT-----CTTACTGAAGAAATGATTGTTTC 465

QY 125 rArgInleu--AenSerIleAenProaenglnValAla-----ArgIleGluVal 141  
 Db 466 ACAGGAACATAGAAATCTTCACTCTCTCTTAACATTCACCTCCCAAAAGCTGCACATC 525

QY 141 leuSerglyAlaThrSerIleTyrglySerglyAlaThrIlyGlyleuIleAsnIleVal 161  
 Db 526 TGTGTCAAAACGAGAGTCTAGCAAAAGAAATAGATGAGATGAGGACAAATGTCCATGCC 585

QY 161 lThrIlySerleuPheleuGlnIleGlnIlePheGlnThrIleGlyValIleGlySerly 181  
 Db 586 TATCAAGTCCAT-----CAGCACTGTTATAG 612

QY 181 leuSergserIlyIleGlyTyrglnValGlyGlnSerValAlaGlyValSergln-- 200  
 Db 613 TTGTGTCAAGTGAATCATTTATGTATGACACAGTACAAATCATCTGACCAAGATCATAT 672

QY 201 -AenGlyAenValIleuAlaArgleuAspValAspIlyrArgThrIlyGlyAlaPheAs 220  
 Db 673 TGAAGTACAGATCAAGCTCTGAAGAT-----GA 702

QY 220 pAlaAengIlyrArgIleAlaProGluProAlaGln-----ThrAspIleGln--As 237  
 Db 703 GTTGGCGAA-AGAAATATCATGTGTGAACCACTGAAGAGAAACAGAAAGCAAA 761

QY 237 pSerIlySerleuSerValAsnThrAsnValAspIlyr-GlnleuAspAspIlyrGlnAsnI 257  
 Db 762 AGGAAAGACTGAAGGCCAAAGCAAGCAGTGTATCAAGCAGTTAGTCAATGATGAT 821

QY 257 leaAsnleuAlaIleuThrIleTyraenAspIlyrGlnAspThrAspIlyrAlaProAspIlyrG 277  
 Db 822 TT--ATTAAAGAAACTGAAGCCAGCTTAGCCAAATTTGAAACATCAACCAAGACCA 878

QY 277 lyAsnArgleuAlaValleu-----PheGlyGlyIlyrProSerleuAenAlaIlelyG 295  
 Db 879 AGCCTCAGATTAAACGCTCTCTCTCAGCTTGTGAAGAAACCAAGATCAAAACCCCTCAC 938

QY 295 lyLeu-----SerleuSergIlyGlnProIlyThr 305  
 Db 939 CACTACACAGATCTGAACCGCAAGAAATTTGAATCACTAACAGAGTCAGAAAGCTTCA 998

QY 305 hlylSerThrPheAenIleAenTyrlleIleAsp-----AspIleuTrpGlyAenThrIle 324  
 Db 999 GAGATCCCTGAGAGCTATGTCTGAACATGTGATGACTTCACTGTGCTGTGAGAGAT 1058

QY 324 snThr-----AenAlaTyTyTyTyArgGlyIlyr-----GlyArgP 336  
 Db 1059 CAGTATCAGAAAGCTTTATTCGTATGCAAAAGAGATAAATGAGACAGCTCGAA 1118

QY 336 heTyrProPheValAlaProPheSerIleAlaIlyr----- 347

```

Db 1119 CAGAGATTTTCAGACCCCATCTCCAGTTCTCAGATCATCAAGGAAATCAGAGAAGAT 1178
Qy 348 -----AlaLeuProIleLeuGlnSerMetAsnLeuProS 359
Db 1179 CTGGAGATTTCTAGAAAATGTACCTGCATACATCTCTCAAGAAATTAATGCCACTA 1238
Qy 359 erAlaThrLeuAspAlaThrLysAlaProGlnAlaArgAlaThrGlyValLeuGlnS 379
Db 1239 GTAGAATTTCTGAT-----ATGTCAGATGCACAAGGTTGGAG 1274
Qy 379 erGluSerLysAlaGluValLeuGlyArgValProAsnLeuAsnLysProLysArg 397
Db 1275 AATCTAGTAAATAAATCAGAAATAAAGAAATAGAGTATACAAAATAAATAAATAA 1330

RESULT 18
BC010844 6042 bp mRNA linear HTC 07-AUG-2002
LOCUS Mus musculus, clone IMAGE:3490875, mRNA.
DEFINITION BC010844
ACCESSION BC010844.1 GI:14789986
VERSION
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Strausberg, R.
Direct Submission
Submitted (12-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www.shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 8 Row: i Column: 9
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis
This clone has the following problem: frame shifted.

FEATURES
Location/Qualifiers
1..6042
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="C57BL/6J"
/clone="IMAGE:3490875"
/tissue_type="mammary tumor. WAP-TGF alpha model. 7 months
old, gross tissue."
/clone_lib="NCI_CGAP_Mam5"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
BASE COUNT 1566 a 1460 c 1764 g 1252 t
ORIGIN

Alignment Scores:
Pred. No.: 10.2 Length: 6042
Score: 102.00 Matches: 57
Percent Similarity: 41.77% Conservative: 47
Best Local Similarity: 22.89% Mismatches: 84
Query Match: 4.98% Indels: 62
DB: 11 Gaps: 11

```

```

US-09-889-746-2 (1-400) x BC010844 (1-6042)
Qy 18 AlaValThrGlnGlnLeuTyAlaGlnProAsnGluSerLeuProThrValGlnLeuGlu 37
Db 253 GCAAAACCAAAAAAATTTGAAAATTTGAAAAGAAAATCCATCAAAATCTATAAAGAGAAG 312
Qy 38 ProValValIleThrIleAspLys-----SerGlyMet----- 48
Db 313 TTTGAAATCTTTAGTCTTGGTCCCTATGTGAGGGGATGCGGATTTTGGGCTGTGTGAAG 372
Qy 49 -----AlaLeuAlaAsnArgIleThrGlnMetProHis 59
Db 373 GAGGTGAGTGAAGCTGGAGCTGGTGTGAGCTTCCCAAGGCGCTCCAAGGGTTTGTGCA 432
Qy 60 ThrThrLysVal-----IleTyrgluGlnIleGlnGlnAlaThrGlySerArg 77
Db 433 GTGCGGAAGTCTGTGAGCGCTACATCAGAAGCTGAATGAGCAGGTGGCACAAGAAGAA 492
Qy 78 GlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSerSerGlyThrThr 97
Db 493 CCTCTGGAGGACCTGCTCCGC-----CTGCTGAGCTC----- 525
Qy 98 SerAsnPheGlyGlnThrMetHisGlyArgGlnValGlnPheLeuLeuAsnGlyValPro 117
Db 526 -----TTCTCCCTCGAATGTTGGTGGTGTGTGTGAGCAGCCTGGAT 570
Qy 118 LeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAsnProAsnGlnValAla 137
Db 571 GTCAAGAGAGCGGCAAGAGAGCGCTCAAGCTT---TCTGTGAACCCCAAAAGTGTCAAC 627
Qy 138 ArgIleGluValLeuSerGlyAlaThrSerIleTyrglySerGlyAlaThrGlyGlyLeu 157
Db 628 AAA-----GTCTGAGTGTGAGGCGCTGAGGCGCTGCTGCTGACAGGC----- 675
Qy 158 IleAsnIleValThrLysSerAspLeuGluGlnGlnPheGluThrArgIleGlyVal 177
Db 676 -----ACAGTGTCAAGCGCTGGAAGACCATGCTACCTAGTGGACATTTGGTGT 723
Qy 178 HisGlySerLysLeu-----SerSer 184
Db 724 GGTGGGACCAAGAGCTTTCTGTCTTAAAGAGGACCAAGAGTACATCCGCGAGAAGAAC 783
Qy 185 GluGlyIleGlyTyrglnValGlyGln-----SerValAlaGlyValSerGlu 200
Db 784 AAAGGTGTCTAAGTTTCAAAAGTCGCGCAGTACCTGACCTGTGTGGTTGAAGAGGTGAAAAGC 843
Qy 201 AsnGlyAsnValLeuAlaArgLeuAspValAspTyArgThrThrGlyGlyAlaPheAsp 220
Db 844 AACGGAGAGATTGTGAGT---CTCTCTGTAGAGCACTCGAGGTTTCTCTGCTTCGCA 900
Qy 221 AlaAsnGlyLysArgIleAlaProGlu 229
Db 901 ACTGA-GGAGCAAGCTGGAACCTGAA 926

RESULT 19
BH368633 634 bp DNA linear GSS 10-DEC-2001
LOCUS AG-ND-163J4.TF ND-TAM Anopheles gambiae genomic clone AG-ND-163J4,
DEFINITION DNA sequence.
ACCESSION BH368633
VERSION BH368633.1 GI:17314677
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 634)
AUTHORS Shetty, J., Malek, J., Koo, H., Collins, F., Gardner, M. and Loftus, B.J.
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)
COMMENT Other_GSSs: AG-ND-163J4.TR

```



```

...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...   ...
262 AAGGATGGCTAGCAAGTCCCTGACGAGGTTTGAAG---GAGATCGTGGCTGGC--- 315
Qy 146 ThrSerIleTyrGlySerGlyAlaThrGlyLeuIleAsnIleValThrLysSerAsp 165
Db 316 -----GGAGGC-----GATCAAGGCGAGGAGTCAA 342
Qy 166 LeuGluGlu-----GluGlnPheGluThrArgIleGlyValHisGlySerLysLeu 182
Db 343 CTTGAAGAGATGACTTTGGAGGATTTTTCAGCAAGAGCTGGTCCGTTAGGAAGAGGAC 402
Qy 183 SerSerGluGlyIleGlyTyrGlnValGlyGlnSerVal-----AlaGly 197
Db 403 GTCACCGGGGAGTTCCTAATCAGGTGGGTGCAGGTTTATTACGTGAGGCTGCCGAT 462
Qy 198 ValSerGluAsnGlyAsnValLeuAlaArgLeu-----AspValAspTyrArg 213
Db 463 ATCAAGCGCGGTGAAGTCAGTATCGACGTTTGGAAACAAAGGTGGCTTAATCATCAG 522
Qy 214 -----ThrThrGlyGlyAlaPheAspAlaAsnGlyLysArgIleAla 227
Db 523 AAGTTGGTGGTCACGCGCAACTGATGAGGC-----GGCAGAGGTAACGACGCGC 573
Qy 228 ProGluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsnThrAsnVal 247
Db 574 GAGGAGGAACCTCTTTGGTAAAGCG-----ACACAGCAG 609
Qy 248 AspTrpGlnLeuAspAspLysGln 255
Db 610 GAAACAGAGGAGGATGATCAAGAA 633

RESULT 21
LOCUS AY105884 1240 bp mRNA linear HTC 25-MAY-2002
DEFINITION Zea mays PC0074613 mRNA sequence.
ACCESSION AY105884
VERSION AY105884.1 GI:21208962
KEYWORDS HTC.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
Clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 1240)
Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whiteitt, M.S.,
Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
Unpublished (2002)
2 (bases 1 to 1240)
Coe, E.C.
Direct Submission
Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
FEATURES
source
1..1240
/organism="Zea mays"
/db_xref="MaizeDB:634118"
/db_xref="taxon:4577"
/clone="PC0074613"
/clone_lib="Maize Mapping Project/DuPont Consensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"
BASE COUNT 358 a 268 c 289 g 323 t 2 others
ORIGIN
Alignment Scores: 0.862 Length: 1240
Pred. No.:

```

```

Score: 101.50 Matches: 90
Percent Similarity: 35.26% Conservative: 50
Best Local Similarity: 22.67% Mismatches: 126
Query Match: 4.96% Indels: 131
DB: 11 Gaps: 22

US-09-889-746-2 (1-400) x AV105884 (1-1240)

Qy 13 ProLeuLeuSerValAlaValThrGln---GlnLeuTyrAlaGlnProAsnGluSerLeu 31
Db 23 CGGTGTGTACCTCGACCTGACGAGCTACCACTGTCAC---GATCTA 67
Qy 32 ProThrValGluLeuGluProValIleThrIleAspLysSerGlyMetAlaLeuAla 51
Db 68 AGCGAGGTGGAGATCCCAACCGCTGGAGGAGTGCACCTCACGCCCAACGCCCTCTCT 127
Qy 52 Asn-----ArgIleThrGlnMetProHisThrLysValIleTyrGluGln 68
Db 128 TCTGTTGATCTCGCATCGGCCAACTCCCCCATCTCCCGCAAGCTCTCATTCGCCGAGAAC 187
Qy 69 Ile-----GlnGluGlnAlaThrGlySerArg 77
Db 188 CTCCTCGACGATGACGCGCTCGGCCACTCAGCTCTCGGAGGCTATTGCCGGTTGAAG 247
Qy 78 Gln-----LeuAlaAspValMetAlaGlnLeuIleProSerLeuGly----- 91
Db 248 GAGTTGCTCTTAGAGATAACAACTTACAGGATTCCTGATGTGGCATATTTCAGGAT 307
Qy 92 -----ValSerSerGlyThrThrSerAsnPheGlyGlnThrMetHisGly 106
Db 308 CTTCTGGTGTTCGATGTATCTTCAATGAGATATCTCT----- 346
Qy 107 ArgGlnValGlnPheLeuLeuAsnGlyValProLeuThrGlySerArgAspLysSerArg 126
Db 347 -----TTGACTGGTCTGTCCCAAGGTTTCCATC 373
Qy 127 GlnLeuAsnSer-----IleAsnProAsnGlnValAlaArgIleGluValLeuSerGly 144
Db 374 AAATTGAAGGAACCTTATGTGTTTCAAAAATGAAGTTGCAAGATGGAAGAGCTGAACAT 433
Qy 145 AlaThrSerIle-----TyrGlySerGlyAlaThrGlyGlyLeuIleAsnIle 160
Db 434 TTCCATGGCTGGAAATCTTGAACCTTGTAGCAACAGATTACGGGTAATGAAATCTTT 493
Qy 161 ValThrLysSerAspLeuGlu----- 168
Db 494 GAAACCTTGACAAATCTACAGAACTGTGGTGGGAAGAAATCGAATTCGAGCTGTCAAC 553
Qy 169 -----GluGlnPheGluThrArgIleGlyVal 177
Db 554 TTGTGTGGTTTGAAGCTAATTAAAGAAATAAGTTTGCAGCAACAGAGCTGATCAATG 613
Qy 178 HisGlySerLys-----LeuSerSerGluGlyIleGly 188
Db 614 CATGGTTTCAGGAATGTGCTCATTAGAGGAGTTATACCTCAGCCATAATGGAATC--- 670
Qy 189 TyrGlnValGlyGlnSerValAlaGlyValSerGluAsnGlyAsnValLeuAlaArgLeu 208
Db 671 -----CAAAAGATGGAAGGCTCTCTACTTGCATAAAC---CTCCGCAATTTTG 715
Qy 209 AspValAspTyrArgThrThrGlyAlaPheAspAlaAsnGlyLysArgIleAlaPro 228
Db 716 GATGTT-----TCAGCTAAC---AAGTACAGCATATA 745
Qy 229 GluProAlaGlnThrAspLysGlnAspSerLysSerValAsnThrAsnValAsp 248
Db 746 GAAATATCGAAACCTTAACACAG---CTGGAGGACCTCTGTTGATGATCACAAC----- 796
Qy 249 TrpGlnLeuAspAspLysGlnAsnIleAsnLeuAlaLeuThrHisTyrAsnAspLysGln 268
Db 797 ---CAAAATACCATCTCTAGAGGCTATAGAGACAGAGCTTTGGCTGGTTCGGCGGAGAAAGTTG 853
Qy 269 AspThrAspTyr-----AlaProAspTyrGlyAsnArgLeu 280

```

```

Db      854  ACCGACATATTCGACGAAATCCCTGTGCAAAAATTCACATTAATCTGACGTTG 913
               |||
Oy      281  AlavalLeuphegylglulysProSerLeuAsnAlaleleSglyseuSerleu----- 298
               |||
Db      914  AGGAAATATTT-----CCAATCTCGAACCAATGATTCAGATATCATGACATGA 964
               |||
Oy      299  -----SerGlnInProlysthr---ThylSeSerThrPheAsnIleAsnIleAsnIle 314
               |||
Db      965  TAACTAGTAATCTCTCCAGATACATAAGCAAAAGAGCTTTGGGGGTCCCC----- 1018
               |||
Oy      315  HlAaPAspLeuTrpGlyAsnThrIleAsnThrAsnAlaIleTyTrpArgArg 331
               |||
Db      1019  -----CGAAGATCATTAACAGCAATCACTGTGTGAGCGCA 1054
               |||
RESULT 22
A2533394      856 bp  DNA  linear  GSS 03-NOV-2000
LOCUS      A2533394
DEFINITION  ENTMB66TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
GENOMIC, DNA sequence.
ACCESSION  A2533394
VERSION    A2533394.1 GI:11089038
KEYWORDS
SOURCE
ORGANISM   Entamoeba histolytica.
            Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE  1 (bases 1 to 856)
            Loftus, B., Van Aken, S. and Fraser, C.
            Determination of clone end sequences from Entamoeba histolytica
            HMI:IMS sheared DNA library
            Unpublished (2000)
JOURNAL
COMMENT    Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: b.loftus@igir.org
            Clones are derived from the Entamoeba histolytica HMI:IMS sheared
            DNA library
            Seq primer: M3-Forward
            Class: shotgun
            High quality sequence start: 25
            High quality sequence stop: 852.
            Location/Qualifiers
                1..856
                /organism="Entamoeba histolytica"
                /strain="HMI:IMS"
                /db_xref="taxon:5759"
                /clone_lib="Entamoeba histolytica Sheared DNA"
                /note="Vector: PHOS1; Site 1: Bat I; Constructed at The
                Institute for Genomic Research (TIGR), Rockville, MD.
                Genomic DNA isolated from broth cultures of E. histolytica
                using a method described by Clark and Diamond (Clark,
                C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
                method for isolate identification. Exp. Parasitol.
                77:450.). The DNA was mechanically sheared to give a
                tight size distribution (~2 kb). The v + i method used for
                the library construction is described in detail in Smith,
                H.O. and Venter, J.C. (Making small insert libraries for
                whole genome shotgun sequencing projects. In Genome
                Sequencing: A Practical Approach, eds. M. Vaundin and B.
                Barrell, Oxford University Press, 1999)."
```

```

US-09-889-746-2 (1-400) x A2533394 (1-856)
Oy      58  ProIeThrThylValIleTyrgluglnIleGlnIleAlaIleThrGlySerArg 77
               |||
Db      43  CCAAGATCTACTGTTACACATCCAGATCAAAAGTTAACCAAAACCCACAGTGA--- 99
               |||
Oy      78  GlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSerSerGlyThr--- 96
               |||
Db      100  -----TACGAAACATCATGTCAGATTAATGACAAAGTACGACCACTCCA 144
               |||
Oy      97  -----ThrSerAsnPhelGlyGlnThr--- 103
               |||
Db      145  GCAATACCGAAGTTCAACAAATTCAGCCAGCTCAACGTCTCAATTAACAAACAGAC 204
               |||
Oy      104  ---MetHlaGlyArgGlnValGlnPheLeuLeuAsnGlyValProLeuThrGlySer--- 121
               |||
Db      205  TCATCAATGGAACAAATGACAGATTAATAACAGGGAGATGCTCTGAAGATCTAGGG 264
               |||
Oy      122  -----ArgAspIleSerArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArg 138
               |||
Db      265  AAAAGTCCAAACCAACCAATGACAGAGAGATTAATCAACAAATGA----- 315
               |||
Oy      139  IleGluValLeuSerGly-----AlaThr 146
               |||
Db      316  ---CAACACCAACTGGGTCAAGAAAAGATAGTGAAGGGAATCTCAACAGCTACC 372
               |||
Oy      147  SerIleTyrglySerGlyAlaIleThrIle-----GlyLeuIleAsnIleValThr--- 162
               |||
Db      373  ACACAGATGACACAGAACCAACAGATAGACCAACAGGTGAAGAAATTAATTAACCCCA 432
               |||
Oy      163  -----LysSerAspLeuGlnIleGlnIlePheGlnIleThrArgIleGlyValHisGly 179
               |||
Db      433  ACAATGCAAAAACAGATGTACCAACAGTGAACATCTACTTAAGAAAGATGAAAAGT 492
               |||
Oy      180  SerLysLeuSer-----SerGlnGlyIleGlyTyrglnValGlyGln--- 193
               |||
Db      493  ACTAAAGCTGAAGAAAGTGAAGAAAGTACAGATGACAGAAAGAAATCTGCACAAAA 552
               |||
Oy      194  -----SerValAlaGlyValSerGlnAsnGlyValLeu 205
               |||
Db      553  CCAAGATGGGGAAGATGTAACAAACAAAGGACAAATGAAAT----- 600
               |||
Oy      206  AlaArgLeuAspValAspTyrgIleThr-----ThrGlyValAlaPheAsp 220
               |||
Db      601  -----ACACAGATAGCAAACTGATTCAAAAGAAAATACAAAACCTAACAGTGA 651
               |||
Oy      221  AlaAsnGlyLysArgIleAlaProGlnProAlaGlnIleAspLysGlnAspSerLysSer 240
               |||
Db      652  CAAATGGAAGAAATAGTTCAAAATGAAAACAAACAACTTCAGAGAA---AAGCAT 708
               |||
Oy      241  LeuSerValSerThrAsnValAspTrpGlnLeuAspAspLysGlnAsnIleAsnLeuAla 260
               |||
Db      709  AATACACCAAAATTAATTCGAAAATTAACACGATGATTAATCAAT----- 756
               |||
Oy      261  LeuThrHisTyAspAspLysGlnAspThrAspTyrgIleAlaProAspTyrgLysAsnArgLeu 280
               |||
Db      757  -----AATATCAATGAGAACAAATGACATCATTAATGATTAATGTCTCAACTATT 810
               |||
Oy      281  AlavalLeu 283
               |||
Db      811  TTCATCTTT 819
               |||
RESULT 23
B0682892
LOCUS      B0682892      791 bp  mRNA  linear  EST 15-JUL-2002
DEFINITION  AGENCOURT 8211651 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6258423
5', mRNA sequence.
ACCESSION  B0682892
VERSION    B0682892.1 GI:21795571
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE  
1 (bases 1 to 791)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: c9apbs-remail.nih.gov  
Tissue Procurement: DCTD/DTF  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLCW2414 row: e column: 16  
High quality sequence stop: 600.

#### FEATURES

source  
1..791  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6258423"  
/clone\_lib="NIH MGC 112"  
/tissue\_type="melanotic melanoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGCAG(G). Library constructed by Ling Hong in the  
laboratory of Gerald M. Rubin (University of California,  
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
Superscript II RT (Life Technologies). Note: this is a  
NIH MGC Library."

BASE COUNT 197 a 275 c 204 g 114 t 1 others

#### ORIGIN

Alignment Scores:  
Pred. No.: 0.924 Length: 791  
Score: 98.50 Matches: 57  
Percent Similarity: 40.00% Conservative: 43  
Best Local Similarity: 22.80% Mismatches: 104  
Query Match: 4.81% Indels: 47  
DB: 14 Gaps: 10

US-09-889-746-2 (1-400) x BQ682892 (1-791)

Qy 5 HisTyrPhe-GlnTrpLeuSerLeuProLeuLeuSerValAlaValThrGlnGlnLeuTy 24  
Db 19 TACTATTTCAGCTGGCTCAGCAACCCACCGCTGTCAGTCCAG-GGCCAGCAGCAGG 77  
Qy 24 rAlaGlnProAsnGlnSerLeuProThrValGluLeuGluProValValThrIleAs 44  
Db 78 CCAGCAGACCCAGCTCCAGCAGCACCACCTCCAGCTGGCGCAGATCATCATGCACAGCC 137  
Qy 44 pIysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrThrIysValI 64  
Db 138 TCAGCAGGCGCAGACC-----ACACCTGTGACAATGAGGTTGG 176  
Qy 64 eTyrGluGluGlnIleGln-----GluGlnAla-- 73  
Db 177 AGAAGTCAGCAGTCAGATTCAGGCTCAGCCAGGGTCAAGCCCAAGCCGCCCA 236  
Qy 74 -ThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSe 93  
Db 237 GAGTGCACCTGGACACACCATGCAGGTGATGCAGCAGATCATC-----ACTAA 284  
Qy 93 rSerGlyThrThrSerAsnPheGlyGlnThrMetHisGlyArgGlnValGlnPheLeu 113  
Db 285 CACAGCAGAGATCCAGCAGATCCCGTGCAGCTGAATCGCGGCAGCTGCAGTATATCCG 344  
Qy 113 uAsnGlyValProLeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIle----- 131  
Db 345 CTTAGCCCGACCTGTATCAGGCACCTCAAGTTGTGCGAGGACAGATCCAGACACTTCCAC 404

Qy 132 -AsnProAsnGlnValAlaArgIleGluValLeuSerGlyAlaThrSerIleTyrGlySe 151  
Db 405 CAATGCTCAACAGATTACACAGACAGAGTCCAGCAAGCAGCAGCAGTTC-----AG 458  
Qy 151 rGlyAlaThrGlyGly-----LeuIleAsnIleValThrLySe 164  
Db 459 CCAGTTCAACATGGACAGCAGAGCTCTACCAGATCCAGCAATGCTCGCTGGGGCCA 518  
Qy 164 rAspLeuGluGluGlnPheGluThrArgIleGlyValHisGlySerLyLeuSerSe 184  
Db 519 GGACCTCGCCAGCCCATGTTTC-----ATCCAGTCAGCCCAACACCCCTC 563  
Qy 184 rGluGlyIleGlyTyrGlnVal-----GlyGlnSerValAlaGlyValSerGl 200  
Db 564 CGACGGCAGGCCCCCAGGTGACCGGCGACTGAGCTGAGCTGGCAAGGCCAAGGAC 623  
Qy 200 uAsnGlyAsnValLeuAlaArgLeuAspValAspTyrArgThrThrGlyGlyAlaPheAs 220  
Db 624 ACCCAACAACATTTTGGCC-----ATACAGCCCCAGGCAATGGGCACAGCCTTCCC 674  
Qy 220 pAlaAsnGlyLyAsArgIleAlaProGlu 229  
Db 675 TCCCCCAGAGGCCCGCGGCGAACCTCAG 702  
RESULT 24  
CNS000W7R  
LOCUS  
DEFINITION Arabidopsis thaliana genome survey sequence SP6 end of BAC T11D13  
of TAMU library from strain Columbia of Arabidopsis thaliana,  
genomic survey sequence.  
ACCESSION AL093237 GI:5294391  
VERSION AL093237.1  
KEYWORDS GSS.  
SOURCE Arabidopsis thaliana.  
ORGANISM Arabidopsis thaliana.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
REFERENCE 1 (bases 1 to 526)  
AUTHORS Salanoubat,M., Choisne,N., Artiguenave,F., Brottier,P., Wincker,P.,  
Samson,D., Saurin,W., Weissenbach,J. and Quetier,P.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 526)  
AUTHORS Genoscope.  
TITLE Direct Submission  
JOURNAL Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : [secref@genoscope.cns.fr](mailto:secref@genoscope.cns.fr))  
- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)  
FEATURES  
source  
1..526  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/strain="Columbia"  
/db\_xref="taxon:3702"  
/clone="T11D13"  
/clone\_lib="TAMU"  
/note="end : SP6"  
BASE COUNT 122 a 144 c 146 g 114 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 0.54 Length: 526  
Score: 98.00 Matches: 27  
Percent Similarity: 46.09% Conservative: 26  
Best Local Similarity: 23.48% Mismatches: 44  
Query Match: 4.79% Indels: 18  
DB: 17 Gaps: 3  
US-09-889-746-2 (1-400) x CNS00W7R (1-526)  
Qy 15 LeuSerValAlaValThrGlnGlnLeuTyAlaGlnProAsnGlnSerLeuProThrVal 34  
Db 190 CTGAGCCCTTCGCTTCGATGATGGTTAT-----AAGGAGGAGAGTGGTACGGTA 240

QY 35 GIUleuGluPro---ValVallierThrlleap----- 44  
 DB 241 CAGCTGACGACCGCGGCAACCGTTACCTGCTGACCTGACCTGAGACCTTGGCCGTTGCA 300  
 QY 45 -----LysSerGlyMetAlaLeuAlaAspArgIleThrGlnMetProHis 59  
 DB 301 TTGGGGGAAGTGTGGTATCTGTGCGACCGCTAAGCCGAGATGTGTCACATGCGCGTGC 360  
 QY 60 ThrThrLysValIleTyrrGluGluGlnIleGlnGluGlnAlaThrGlySerArgGlnLeu 79  
 DB 361 TCCGTAACCTCATCTCTCCCGGAAACATCGAGAACAGACCTGCTGTGATCACCGACGTA 420  
 QY 80 AlaAspValMetAlaGlnLeuIleProSerLeuGlyValSerSerGlyThrThrSerAen 99  
 DB 421 ACAGAGATACTGGCCAAACGAGTGGCCGCGGCTGGCCCTTCGCGCAACGACGACCAAC 480  
 QY 100 PheGlyGlnThrMetHisGlyArgGlnValGlnPheLeuLeuAen 114  
 DB 481 GTGGGGCAATCGCTGCGCGCGCTCTGTGCTTATCATATCATGAC 525

RESULT 25  
 B0837115 593 bp mRNA linear EST 08-AUG-2002  
 LOCUS r138d04.y1 Meloidogyne hapla J2 PAMP1 v1 Meloidogyne hapla cDNA 5'  
 DEFINITION similar to TR:Q9Y419 Q9Y419 TYROSINE PHOSPHATASE IA-2BETA //, mRNA  
 sequence.

ACCESSION B0837115  
 VERSION B0837115  
 KEYWORDS EST  
 SOURCE Meloidogyne hapla.  
 ORGANISM Meloidogyne hapla.

REFERENCE 1 (bases 1 to 593)  
 AUTHORS McCarter,J., Clifton,S., Chiappelli,B., Page,D., Martin,J., Wylie,T.,  
 Dante,M., Maria,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,  
 Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagaris,I., R.,  
 Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Stepien,  
 M., Allen,M., Person,B., Swaller,T., Harvey,N., Schuck,R., Kohn,S.,  
 Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and  
 Wilson,R.

TITLE The Washington Univ. Nematode EST Project, 1999  
 JOURNAL Unpublished (1999)  
 COMMENT The Washington Univ. Nematode EST Project, 1999  
 Contact: McCarter JP  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.wustl.edu

The library was constructed by Claire Murphy and Dr. James McCarter  
 at Washington University, St. Louis. U2 were provided by Dr.  
 Valerie Williamson of the University of California at Davis  
 (vwilliamson@ucdavis.edu).  
 Seq primer: -40RP from Gibco  
 High quality sequence scop: 423.

# FEATURES

source  
 1..593  
 Location/Qualifiers  
 /organism="Meloidogyne hapla"  
 /db\_xref="taxon:6305"  
 /clone\_lib="Meloidogyne hapla J2 PAMP1 v1"  
 /dev\_stage="J2"  
 /lab\_host="DH10B"  
 /note="Vector: PAMP1 (Gibco); Site 1: NotI; Site 2: SalI;  
 The library was constructed by Claire Murphy and Dr. James  
 McCarter at Washington University, St. Louis. The cDNA was  
 made by using Dynabead oligo-dT priming (Dyna1). PCR based  
 library using a modified protocol from the SMART PCR cDNA  
 Synthesis Kit from Clontech. Directionally cloned into the  
 UNG sites of pAMP1. U2 were provided by Dr. Valerie  
 Williamson of the University of California at Davis  
 (vwilliamson@ucdavis.edu)."

BASE COUNT 186 a 97 c 145 g 165 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0.657 Length: 593  
 Score: 98.00 Matches: 52  
 Percent Similarity: 39.82% Conservative: 36  
 Best Local Similarity: 23.53% Mismatches: 79  
 Query Match: 4.79% Indels: 54  
 DB: 14 Gaps: 9

US-09-889-746-2 (1-400) x B0837115 (1-593)

QY 53 ArgIleThrGlnMetProHisThrThrIleValIleTyrrGluGlnIleGlnGluGln 72  
 DB 39 AAGCGTAAACATATCTCTACACACTTGTGTGATCGAAGCGTCAACGTTCAAGCACTCG 98  
 QY 73 AlatrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyVal 92  
 DB 99 ACTACTTGTGCTATTCACGAAATGAG-----ACTTTGGGGTGG 137  
 QY 93 SerSerGlyThrThrSerAspPheGlyGlnThrMetHisGlyArgGlnValGlnPheLeu 112  
 DB 138 TCCAAATGGGTCTGATATGACATTGGG---ACAGACACGCTCTTTGGCTTTCTGTTT 194  
 QY 113 LeuAsnGlyValProLeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAen 132  
 DB 195 TTACAT-----TTGGAATCTCAACGAAATTAACAAATGGAATCTGTGAAG 245  
 QY 133 -----ProAsnGlnValAlaArgIleGluValLeuSerGlyValThrSerIle 148  
 DB 246 GATTATGAAAGCGCCGAGAGAAAGTGGCTCACTGTAAGCTGTAAGAGAGAGAAATGTT 305  
 QY 149 TyrrGlySerGlyAlaThrGlyGlyLeuIleAenIleValThrLysSerAspLeuGlu 168  
 DB 306 GCCAAATTTGATTAATGAATGAACTCTT----- 332  
 QY 169 GluGlnPheGluThrArgIleGlyValHisGlySerLysLeuSerSerGluIleGly 188  
 DB 333 CCATATGATGATTCGCGTGTGCACTTCAAGCAATTTGGCAGATTTGGTGATGA 392  
 QY 189 TyrrGlnValGlyLysSerValAlaGlyValSerGluAsnGlyAsnValLeuAlaArgLeu 208  
 DB 393 -----AAGGTTATATTAATCTAGTATG 416  
 QY 209 AspValAspTyrrArgThrThrGlyGlyAlaPheAspAlaAsnGlyValArgIleAsp 228  
 DB 417 -----ATTTTGTGATTCGAATCCAGAACACCGGCTTAT 449  
 QY 229 GluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsnThrAsnValAsp 248  
 DB 450 ATTGGCGGCCCAACACCTCTAGACGAACTGTTCTCAATT----- 491  
 QY 249 TrpGlnLeuAspAspLysGlnAsnIleAsnLeuAlaLeuThrHisTyrrAsnAspLysGln 268  
 DB 492 TGGCAATGATTTGGGAACAGGAATTTGGGTTA---ATCCGCATTGTGATGAAGA 548  
 QY 269 Asp 269  
 DB 549 GAC 551

RESULT 26  
 B035098 615 bp mRNA linear EST 07-DEC-2001  
 LOCUS B035098 NIBB Mochi1 normalized Xenopus neurula library Xenopus  
 DEFINITION laevis cDNA clone X1031p06 5', mRNA sequence.  
 ACCESSION B035098  
 VERSION B035098  
 KEYWORDS EST  
 SOURCE African clawed frog.  
 ORGANISM Xenopus laevis  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;

```

Xenopodinae; Xenopus.
1 (bases 1 to 615)
REFERENCE
AUTHORS Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara
,Y.
TITLE Expressed genes in X. laevis embryo
JOURNAL Unpublished (2001)
COMMENT Contact: Tadaseu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
FEATURES
Location/Qualifiers
1..615
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="XLO31p06"
/clone_lib="NIBB Mochii normalized Xenopus neurula
library"
/tissue_type="whole embryo"
/dev_stage="stage 15"
/notes="Vector: DBSRN3; Site 1: NotI; Site 2: EcoRI; cDNAs
were oligo-dT primed and directionally cloned. Staging
according to Nieuwkoop and Faber. Library is subtracted
and was constructed by N. Garrett and A.M. Zorn,
(Wellcome/CRC Institute)."
BASE COUNT 166 a 159 c 162 g 128 t
ORIGIN
Alignment Scores:
Pred. No.: 0.915 Length: 615
Score: 97.00 Matches: 55
Percent Similarity: 35.25% Conservative: 31
Best Local Similarity: 22.54% Mismatches: 78
Query Match: 4.74% Indels: 80
DB: 13 Gaps: 10
US-09-889-746-2 (1-400) x BJ035098 (1-615)
Qy 112 LeuLeuAsnGlyValProLeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIle 131
Db 2 GTTCTGTGGGTCTACTCTGGATCCAGAGAAATGTCAGAGACGCGCTTCAAAACA 61
Qy 132 AsnProAsnGlnValAlaArgIleGluValLeuSerGlyAlaThrSerIleTyrGlySer 151
Db 62 ATCCGCAGCTTCTCGATAAATAGACAGATATCCGGGATCCATCA----- 109
Qy 152 GlyAlaThrGlyGlyLeuIleAsnIleValThrLysSerAspLeuGluGluGlnPhe 171
Db 110 -----CAGCTGGCAGAACTAGAAAAGGAT----- 133
Qy 172 GluThrArgIleGlyValHisGlySerLysLeuSerSerGluGlyIleGlyTyrGlnVal 191
Db 134 -----GTCCACACAGCTTCAGCAAGCCCAAGTGTGTGGGAGGATGGCA 178
Qy 192 GlyGlnSerValAlaGlyValSerGluAsnGlyAsnValLeuAlaArgLeuAspValAsp 211
Db 179 GGCTGGGCTGTCACTGGTGTCT-----TCCTTGACCTCCAAGTTC----- 220
Qy 212 TyrArgThrThrGlyGlyAlaPheAspAlaAsnGlyLysArgIleAlaProGluProAla 231
Db 221 ATCCGCACAGGGGTGAGCCAGGATGATCTCTTTCAGAGGGGCGACCACTTCCA 280
Qy 232 GlnThr----- 233
Db 281 GCCACCATAGAACGAAACCGGATGACTGTCTTCACATCTGCTCCCTCGTGGC 340
Qy 234 -----AspLysGlnAspSerLysSerLeuSer 242
Db 341 TCAACCAACCCATCCAGCTATGAACCCGAGAGAAAAGGTGCTTCAGATTCCTTAGAT 400
Qy 243 ValAsnThrAsnValAspTrp-----GlnLeuAspAspLys-----GlnAsn 256

```

```

Db 401 CGATGGGATGATGAAGATTGGGGAGCTTAGAGGACACAGACAGAGAGGTCAGACT 460
Qy 257 IleAsnLeuAlaLeuThrHisTyrAsnAspLysGlnAspThrAspTyrAlaProAspTyr 276
Db 461 GTG-----AATGATGACTGGGACACAGACTGGGGTCAGAGCAAA 499
Qy 277 GlyAsnArgLeuAlaValLeuPheGlyGluLysProSerLeuAsnAlaIleLysGlyLeu 296
Db 500 ACGCAGCAGAAAGACTGTC-----CACTCTTCTCTCAAGTTCCAAAGAGAG 547
Qy 297 SerLeuSerGluGlnProLysThrThrLysSerThrPheAsnIleAsnTyrHisAsp 316
Db 548 CAGGTTTTGCACCTCTCAACAAACTTCTGCTCTC-----GAT 586
Qy 317 AspLeuTrpGly 320
Db 587 GATGCTGGGA 598
RESULT 27
LOCUS BQ748998
DEFINITION BQ748998 812 bp mRNA linear EST 17-JUL-2002
IMAGE:5714661 5', mRNA sequence.
ACCESSION BQ748998
VERSION BQ748998.1 GI:21895785
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 812)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE NIH-MGC National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Straubeberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)
Seq primer: pYX-5.
Location/Qualifiers
1..812
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:5714661"
/clone_lib="NIH BMAP_FBO"
/tissue_type="Whole Brain"
/dev_stage="embryo 12.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/notes="Organ: brain; Vector: pYX-Asc; Site 1: EcoR I;
Site 2: Not I; The library was constructed according to
Ronald, Lennon and Soares, Genome Research 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
, is TGAGAGAGCC. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP): 'Gene Discovery in the Developing Mouse Nervous
System', supported by National Institutes of Mental Health
(NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT 216 a 253 c 207 g 132 t 4 others

```



## ORIGIN

## Alignment Scores:

Pred. No.: 1.45 Length: 812  
 Score: 97.00 Matches: 41  
 Percent Similarity: 45.45% Conservative: 29  
 Best Local Similarity: 26.62% Mismatches: 59  
 Query Match: 4.74% Indels: 26  
 DB: 14 Gaps: 5

US-09-889-746-2 (1-400) x B0748998 (1-812)

QY 5 H1ETYPhe-G1TTPLeuSerLeuProLeuLeuSerValAlaValThrGlnGlnLeuTy 24  
 :::::::::::::::::::::  
 Db 62 TACTACTTACCGCTGGCTCAGCAGCCACCTGCTGTCCAG-GTCCAGGACGACGACG 120  
 QY 24 rLaglnProAenGlnSerLeuProThrValGlnLeuGlnProValAlaIleThrIleAs 44  
 :::::::::::::::::::::  
 Db 121 CCAGCAGACCAACGATTCTACGACCAACCTCGACAGATCATCATTCACAGCC 180  
 QY 44 phySerGlyMetAlaLeuAlaAanArgIleThrGlnMetProH1eThrIleVal-- 63  
 :::::::::::::::::::::  
 Db 181 TCAGCAGGGTCAGACC-----ACACCCGTGACCATGACGTTGG 219  
 QY 64 -----TlEtyGlnGlnGlnIleGlnGlnAla----- 73  
 Db 220 AGAAGGTCAGACGATGTCAGATTGTACAGGCCCAACCTCAGGGTCAGGCCACGACCA 279  
 QY 74 -ThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSe 93  
 :::::::::::::::::::::  
 Db 280 GAGTGGCACTGGACACACCATGACGATGACGAGATCATT-----ACCA 327  
 QY 93 rSerGlyThrThrSerAsnPhenGlyGlnThrMetH1eGlyArgGlnValGlnPheLeu 113  
 :::::::::::::::::::::  
 Db 328 CACAGAGAGATCCACAGATCCCGGTGACGCTGATGCTGGCCACCTGCAATATATCCG 387  
 QY 113 uAenGlyValProLeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIle----- 131  
 :::::::::::::::::::::  
 Db 388 ATTAGCCCAAGCTGTATGAGCCACCAAGTGTCCAAAGACAGATCAGACACTTCTAC 447  
 QY 132 -AsnProAenGlnValAlaArgIleGlnValLeuSerGly 144  
 :::::::::::::::::::::  
 Db 448 CAATGCCCAACAGATCACAAGACAGAGTCCAGCAAGA 487  
 RESULT 28  
 CENS0746Z/c 841 bp DNA linear GSS 07-JUL-2001  
 LOCUS clone BA0AB025G03 of library BA0AB from strain CL1B 210 of  
 DEFINITION Kluyveromyces lactic, genomic survey sequence.  
 ACCESSION AL428433  
 VERSION AL428433.1 GI:12211627  
 KEYWORDS GSS.  
 SOURCE Kluyveromyces lactic.  
 ORGANISM Kluyveromyces lactic.  
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 REFERENCE 1 (bases 1 to 841)  
 AUTHORS Soucier,J.L., Aigle,M., Artiguenave,F., Blandin,G.,  
 Bojotin-Pukhara,M., Bon,E., Brotier,P., Cabaregola,S.,  
 de-Montigny,J., Dujon,B., Durzan,P., Lepingle,A., Llorente,B.,  
 Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,  
 Saurin,M., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,  
 Winkler,P. and Weissenbach,J.  
 TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
 yeast species for molecular evolution studies  
 JOURNAL FEBS Lett. 487 (1), 3-12 (2000).  
 MEDLINE 20584711  
 PUBMED 11152886  
 REFERENCES 2 (bases 1 to 841)  
 AUTHORS Bojotin-Pukhara,M., Toffano-Nioche,C., Artiguenave,F.,  
 Duchateau-Nguyen,G., Lemaire,M., Marelsse,R., Montrocher,R.,  
 Robert,C., Terrier,M., Winkler,P. and Wesolowski-Louvel,M.  
 TITLE Genomic exploration of the hemiascomycetous yeasts: 11.

JOURNAL Kluyveromyces lactic  
 MEDLINE FEBS Lett. 487 (1), 66-70 (2000)  
 PUBMED 20584721  
 11152886  
 REFERENCE 3 (bases 1 to 841)  
 AUTHORS Soucier,J.L., Aigle,M., Artiguenave,F., Blandin,G.,  
 Bojotin-Pukhara,M., Bon,E., Brotier,P., Cabaregola,S.,  
 de-Montigny,J., Dujon,B., Durzan,P., Lepingle,A., Llorente,B.,  
 Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,  
 Saurin,M., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,  
 Winkler,P. and Weissenbach,J.  
 TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of  
 yeast species for molecular evolution studies  
 JOURNAL FEBS Lett. 487 (1), 3-12 (2000).

## COMMENT

This GSS is part of a random genomic sequencing program of thirteen  
 yeast species: *Saccharomyces bayanus* var. *uvatum*, *Saccharomyces*  
*exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*,  
*Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces*  
*lactic* var. *lactic*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia*  
*angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*,  
*Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to  
 5 kb were prepared and both extremities were sequenced. See  
 keywords for description of this sequence and for the sequence of  
 the other extremity of this insert.

## FEATURES

## source

1..841  
 /organism="Kluyveromyces lactic"  
 /strain="CL1B 210"  
 /variety="lactic"  
 /db\_xref="taxon:28985"  
 /clone="BA0AB025G03"  
 /clone\_1lb="BA0AB"  
 complement(2..5796)  
 /note="similar to *Saccharomyces cerevisiae* ORF YJL041w [  
 NSP1 ; nuclear pore protein ]"  
 /evidence="not experimental"

## BASE COUNT

193 a 170 c 207 g 270 t 1 others

## ORIGIN

Alignment Scores:  
 Pred. No.: 1.53 Length: 841  
 Score: 97.00 Matches: 70  
 Percent Similarity: 36.33% Conservative: 35  
 Best Local Similarity: 24.22% Mismatches: 114  
 Query Match: 4.74% Indels: 71  
 DB: 17 Gaps: 15  
 US-09-889-746-2 (1-400) x CENS0746Z (1-841)  
 QY 75 GlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSer 94  
 :::::::::::::::::::::  
 Db 802 GGGTGGCAACAGCTTAAGACACTCA-----CCCAAGATGATTCATTT 755  
 QY 95 GLYThr-----ThrSerAsnPhenGlyGlnThrMet 104  
 :::::::::::::::::::::  
 Db 754 GGTACCAACTCTGAACAAAGACCATGCAAGCCCGCTTTAATTTCCGA-----CTA 701  
 QY 105 H1eGlyArgGlnValGlnPheLeuAenGlyVal-----ProLeuThrGly 120  
 :::::::::::::::::::::  
 Db 700 TCTGGAATTAATGCAATGCTCAACCTCGAGTGCATGCTTAAACCATATTTGGT 641  
 QY 121 SerArgAspIleSerArgGln-----LeuAsnSerIleAsnProAenGlnValAlaArg 138  
 :::::::::::::::::::::  
 Db 640 GCCAATACCAACTTCAACATCAAGCTTCAACACTCCGTTAAACCAAGCAACTCGAAA 581  
 QY 139 lLeGlnValLeuSerGlyValAlaThrSerIleTyGly----- 150  
 :::::::::::::::::::::  
 Db 580 CCTAAT-----ACAAATGCCACATCATCTTTGGTGGGACCAACCATGATACAAACCA 527  
 QY 151 SerGlyAlaThrGlyGlyLeuIleAsnIleValIThrIlySerAspLeuGlnGlnGln 170  
 :::::::::::::::::::::  
 Db 526 GCCGGTGTACTTCAAGGTTCACCTTGTCTGGCGCTGAGAGATCCCAACAGCAACT 467  
 QY 171 PheGlnThrArgIleGlyValH1eGlySerIlyLeuSerSerGlyIlylLeGlyTyGln 190  
 :::::::::::::::::::::  
 Db 466 TTTTGGCTAAGTCTG-----GATACAAACCTGGAAGAAAACCTCAGGCCCAAG 416



```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5504081"
/clone_lib="NCI CGAP Ova44"
/lab_host="DH10B (TI phase-resistant)"
/notes="Organ: ovary, FMS-treated; Vector:
pCMV-Sport6.cdb; Site_1: EcoRV; Site_2: NotI; Cloned
unidirectionally. Primer: Oligo dt. Average insert size
2.2 kb. Library constructed by Life Technologies. Note:
this is a NCI CGAP library."

```

BASE COUNT 301 a 328 c 265 g 218 t 1 others

## ORIGIN

Alignment Scores:

Pred. No.:	2-43	Length:	1113
Score:	97.00	Matches:	41
Percent Similarity:	45.45%	Conservative:	29
Best Local Similarity:	26.62%	Mismatches:	59
Query Match:	4.74%	Indels:	26
DB:	13	Gaps:	5

US-09-889-746-2 (1-400) x BM461330 (1-1113)

```

QY 5 HETYPhe-GlntPLeuSerLeuProLeuLeuSerValAlaValThrgInGlnLeuTy 24
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 281 TACTACTTCAAGCTGCTGCTCAGAGCCAGCTGCTGCTCAG-GTCCAGGAGCAGCAGCAAG 339
    |||
QY 24 rAlaGlnProAnGluSerLeuProThrValGluLeuGluProValValIleThrIleAs 44
    |||
DB 340 CCAGGAGCACCAGGTTCTACGACCAACCACTGAGCAGCATATCATCTTGCACAGCC 399
    |||
QY 44 pLysSerGlyMetAlaLeuAlaAnArgIleThrgInMetProIleThrThryVal-- 63
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 400 TCAGCAGGGTCAGACC-----ACACCCGTGACCATGAGAGTTGG 438
    |||
QY 64 -----IleTyGlnGluGlnIleGlnGlnAla----- 73
    |||
DB 439 AGAAGGTCAGAGGTCAGATTGTACAGCCCACTCAGGCTCAGGCTCAGCAGCAGCA 498
    |||
QY 74 -ThrgIserArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSe 93
    |||
DB 499 GAGTGCACTGAGCAGCAGCATGAGGTCGATGACGAGCATCAT--ACCA 546
    |||
QY 93 rSerGlyThrThrSerAspPheGlyGlnThrMetHISGlyArgGlnValGlnPheLeu 113
    ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
DB 547 CACAGAGAGATCCACAGATCCCGGTGACGCTGAATCTCGGCACGCAATATCCG 606
    |||
QY 113 uAnGlyValProLeuThrgIserArgAspIleSerArgGlnLeuAnSerIle----- 131
    |||
DB 607 ATTAGCCCAAGCTGTATCAGGACCAAGTTGTCCAAAGCAGATCAGACACTTGCTAC 666
    |||
QY 132 -AsnProAnGlnValAlaArgIleGluValLeuSerGly 144
    |||
DB 667 CAATGCCCAAGATCACAACGACAGAGTCCAGCAAGA 706
    |||

```

RESULT 31 AK011327 1482 bp mRNA linear HTC 19-JAN-2002  
LOCUS AK011327  
DEFINITION Mus musculus 10 days embryo whole body cDNA, RIKEN full-length  
enriched library, clone:2610005N20:nuclear transcription factor-Y  
gamma, full insert sequence.

ACCESSION AK011327  
VERSION AK011327.1 GI:12847379  
KEYWORDS HTC, CAP trapper.  
SOURCE Mus musculus (strain: C57BL/6J) 10 days embryo cDNA to mRNA,  
clone:2610005N20.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 Carninci, P. and Hayashizaki, Y.  
AUTHORS High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279623  
PUBMED 10349636  
REFERENCE 2

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL 20499374  
PUBMED 11042159  
REFERENCE 3

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Kono, H., Akiyama, J., Nishi, K., Kitenunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasai, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Obara, E., Watanabe, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multiplexed sequencer  
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL 20530913  
PUBMED 11076861  
REFERENCE 4

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Arakawa, T., Hara, A., Fukunishi, Y., Kono, H., Adachi, J., Fukuda, S.,  
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanka, I.,  
Saito, T., Okazaki, Y., Gojodori, T., Bono, H., Kasukawa, T., Saito, R.,  
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
Kuehl, P., Lewis, S., Matsuo, Y., Nakado, I., Pesole, G.,  
Quackenbush, J., Schriml, L. M., Stubbs, F., Suzuki, R., Tomita, M.,  
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H.,  
Baldarelli, R., Barsn, G., Blake, J., Boffelli, D., Bojunga, N.,  
Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bult, C.,  
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
Hottmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P.,  
Marchionni, L., Mashima, J., Mazzarelli, J., Mombere, P., Nordone, P.,  
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K. P., Suzuki, H.,  
Toyo-Oka, K., Wang, K. H., Weltz, C., Whitteker, C., Wilmink, J.,  
Wynshaw-Borje, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohetsuki, S.  
and Hayashizaki, Y.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409 (6821), 685-690 (2001)

JOURNAL 21085660  
PUBMED 11217851  
REFERENCE 5 (bases 1 to 1482)

Adachi, J., Aizawa, K., Akahira, S., Akiyama, T., Aono, H., Arai, A.,  
Arakawa, T., Baldarelli, R., Bono, H., Brownstein, M., Bult, C.,  
Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hangaki, T.,  
Hara, A., Hayatsu, N., Hill, D., Hiramoto, K., Hirooka, T., Hori, F.,  
Hume, D., Imocani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T.,  
Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koya, S.,  
Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K.,  
Numazaki, R., Ono, M., Okazaki, Y., Okido, T., Owa, C., Quackenbush, J.,  
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,  
Schriml, L., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T.,  
Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F.,  
Tanaka, T., Tajima, Y., Toya, T., Yamamura, T., Yanaka, I.,  
Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and  
Hayashizaki, Y.

Direct Submission  
Submitted (10-JUL-2000) Yoshinori Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Teikoku-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail: genome-res@res.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp, Tel: 81-45-503-9222,  
Fax: 81-45-503-9216)  
Please visit our web site (http://genome.gsc.riken.go.jp/) for  
further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5', GAGAGAGAGAGATCAAGAGCTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 7.5 and subtraction to Rot = 37.5. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCGAGTGAATTAATATCCCTCCCTCCCTCC 3']. cDNA was cleaved with XhoI and SstI. Cloning sites, 5' end: XhoI; 3' end: SstI. Host: SOLR.

FEATURES  
Location/Qualifiers  
1..1482  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="PANTOM DB:2610005N20"  
/db\_xref="MGD:MGI:1909008"  
/db\_xref="taxon:10090"  
/clone="2610005N20"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"  
/dev\_stage="10 days embryo"  
1..1482  
/gene="Nfyc"  
misc\_feature  
1..1482  
/gene="Nfyc"  
/note="data source:MGD, source key:MGI:107901, evidence:ISS  
nuclear transcription factor-Y gamma"  
/db\_xref="MGD:MGI:107901"  
BASE COUNT 382 a 424 c 364 g 312 t  
ORIGIN

Alignment Scores:  
Pred. No.: 3.9 Length: 1482  
Score: 97.00 Matches: 41  
Percent Similarity: 45.45% Conservative: 29  
Best Local Similarity: 26.62% Mismatches: 59  
Query Match: 4.74% Indels: 26  
DB: 11 Gaps: 5

US-09-889-746-2 (1-400) x AK011327 (1-1482)

QY 5 HistyPhe-GlnTrpLeuSerLeuProLeuSerValAlaValThrGlnGlnLeuTy 24  
Db 606 TACTACTTCAGCTGGCTCAGAGCCACTGCTGTCAG-GTCCAGGAGCAGCAGCAAG 664

QY 24 rAlaGlnProAsnGluSerLeuProThrValGluGluProValValThrIleAs 44  
Db 665 CCAGCAGACCACCATGTTCTACGACCACTCCAGCTGGACATCATCATTCACAGCC 724

QY 44 pLysSerGlyMetAlaLeuAlaAsnArgIleThrGlnMetProHisThrIleVal-- 63  
Db 725 TCAGCAGGGTCAGACC-----ACACCCGTGACCATGTCAGGTGG 763

QY 64 -----IleTyGluGluGlnIleGlnGlnAla----- 73  
Db 764 AGAAGGTCAGCAGGTGAGATTGTACAGGCCCACTCAGGTGAGCCCGCAGCAGCCCA 823

QY 74 -ThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValse 93  
Db 824 GAGTGGCACTGGACACCATGAGGTGATGCAGCATCATTT-----ACCAA 871

QY 93 rSerGlyThrThrSerAsnPheGlyGlnThrMetHisGlyArgGlnValGlnPheLeu 113  
Db 872 CACAGGAGAGATCCAGCATCCCGGTGCGAGTGAATGCTGGCCAGCTGCAATATATCCG 931

QY 113 uAsnGlyValProLeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIle----- 131  
Db 932 ATTAGCCAGCCTGTATCAGGACCCCAAGTTGTCCAGGAGCAGATCCAGACACTTGTCTAC 991

QY 132 -AsnProAsnGlnValAlaArgIleGluValLeuSerGly 144  
Db 992 CATATCCCAAGATCATCACAGACAGAGGTCAGCAAGGA 1031

RESULT 32  
BQ835748  
LOCUS  
DEFINITION  
BQ835748 605 bp mRNA linear EST 08-AUG-2002  
rf57C08.y1 Meloidogyne hapla J2 pAMP1 v1 Meloidogyne hapla cDNA 5', similar to TR:Q9Y419 Q9Y419 TYROSINE PHOSPHATASE 1A-2BETA ; mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
BQ835748  
BQ835748.1 GI:22140062  
EST  
Meloidogyne hapla.  
Meloidogyne hapla  
Tylenchoidea; Nematoda; Chromadorea; Tylenchida; Tylenchina; Eukaryota; Metazoa; Heteroderidae; Meloidogyninae; Meloidogyne.  
REFERENCE  
1 (bases 1 to 605)  
AUTHORS  
McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagareishvili, R., Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe, M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.  
The Washington Univ. Nematode EST Project, 1999  
Unpublished (1999)  
Contact: McCarter JP  
The Washington Univ. Nematode EST Project, 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
The library was constructed by Claire Murphy and Dr. James McCarter at Washington University, St. Louis. J2 were provided by Dr. Valerie Williamson of the University of California at Davis (vmwilliamson@ucdavis.edu).  
Seq primer: -40RP from Gibco  
High quality sequence stop: 416.  
Location/Qualifiers  
1..605  
/organism="Meloidogyne hapla"  
/db\_xref="taxon:6305"  
/clone\_lib="Meloidogyne hapla J2 pAMP1 v1"  
/dev\_stage="J2"  
/lab\_host="DH10B"  
/note="Vector: pAMP1 (Gibco); Site 1: NotI; Site 2: SalI;  
The library was constructed by Claire Murphy and Dr. James McCarter at Washington University, St. Louis. The cDNA was made by using Dynabead oligo-dT priming (Dyna). PCR based library using a modified protocol from the SMART PCR cDNA Synthesis Kit from Clontech. Directionally cloned into the UDG sites of pAMP1. J2 were provided by Dr. Valerie Williamson of the University of California at Davis (vmwilliamson@ucdavis.edu)."  
BASE COUNT 194 a 99 c 148 g 164 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.02 Length: 605  
Score: 96.50 Matches: 58  
Percent Similarity: 40.11% Conservative: 38  
Best Local Similarity: 24.27% Mismatches: 84  
Query Match: 4.71% Indels: 59  
DB: 14 Gaps: 12

US-09-889-746-2 (1-400) x BQ835748 (1-605)

QY 53 ArgIleThrGlnMetProHisThrIleValIleTyGluGlnIleGlnGln 72  
Db 39 AGAGCTGAAGAAACATCCCTCACAGATTGCTGTGATCGAAGGTCACAGGTCACAGCTCG 98

Qy 73 AlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyVal 92  
 Db 99 ACTACTTCCTCATCCACGAAATGAG-----ACTTTGGGGGTG 137  
 Qy 93 SerSerGlyThrThrsSeranPheGlyInThrMetHisGlyArgGlnValGlnPheLeu 112  
 Db 138 TCCAAATGGGTCTGATATGACATTGGG---ACAGGACAGCC---CTTTGGCTTTTCG 191  
 Qy 113 LeuAnGlyValProLeuThrGlySerArgAspIleSerArgGlnLeuAnSerIleAn 132  
 Db 192 CAAAA-----CATTTGACATCCAAACGAATTAACAACAAATGGGAATCTGTGAAG 245  
 Qy 133 -----ProAnGlnValAlaArgIleGluValLeuSerGlyAlaThrSerIle 148  
 Db 246 GATTATGAAGCCCGGAGAGAGAGTGGCTCAGTAAAGTGGCTAGAGAGAGAGAGATGTT 305  
 Qy 149 TyrGlySerGlyAlaThrGlyLeuIleAnIleValThrLysSerAspLeuGlu 168  
 Db 306 GCCAAAAATTTGATATGAACTCTT----- 332  
 Qy 169 GlnGlnPheGluThrArgIleGlyValHisGlySerLysLeuSerSerGlyIleGly 188  
 Db 333 CCATATGATGAATCCGCGTGTGCAACTTCAATGAAGGATTTGGCAGAAATGGGTATGA 392  
 Qy 189 TyrGlnValGlyGlnSerValAlaGlyValSerGluAnGlyAsnValLeuAlaArgLeu 208  
 Db 393 -----AAGGTTATATATCTACTATGCTACTATG 416  
 Qy 209 AspValAspTyrArgThrThrThrGlyAlaPheAspAlaAsnGlyValArgIleAlaPro 228  
 Db 417 -----ATTTGATTCGAATCCAAACACCGGCTTAT 449  
 Qy 229 GluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsnThrAsnValAsp 248  
 Db 450 ATTACCGGCCCAACACCTCTAGACGAACACTGCTTCATTT----- 491  
 Qy 249 TrpGlnLeuAspAspLysGlnAsnIleAnLeuAlaLeuThrHisTyrAsnAspLysGln 268  
 Db 492 TCGCAATGATATTGGGACAGGGAATTTGGTTA---ATCGTCAATTTGCTGTAAGA 548  
 Qy 269 AspThr-----AspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeuPheGly 285  
 Db 549 GCAATTAATCTTAATAATATGGCCAGAGAGAGATGAAG-----CTGTTTGT 596  
 RESULT 33  
 B0836176 605 bp mRNA linear EST 08-AUG-2002  
 LOCUS r142g10.y1 Meloidogyne hapla J2 PAMP1 v1 Meloidogyne hapla cDNA 5',  
 DEFINITION similar to TR:Q9Y419 Q9Y419 TYROSINE PHOSPHATASE IA-2BETA ;, mRNA  
 sequence.  
 ACCESSION B0836176  
 VERSION B0836176.1 GI:22140490  
 SOURCE EST.  
 ORGANISM Meloidogyne hapla.  
 Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
 Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.  
 1 (bases 1 to 605)  
 McCarter, J., Clifton, S., Chiappelli, B., Pape, D., Martin, J., Wylie, T.,  
 Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,  
 Gibbons, M., Richter, B., Bennett, J., Franklin, C., Tsagarisavlil, R.,  
 Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe,  
 M., Allen, M., Person, B., Swaller, T., Harvey, N., Schuck, R., Kohn, S.,  
 Shih, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and  
 Wilson, R.  
 The Washington Univ. Nematode EST Project, 1999  
 Unpublished (1999)  
 CONTACT McCarter JP  
 The Washington Univ. Nematode EST Project, 1999  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
 Tel: 314 286 1800  
 Fax: 314 286 1810

Email: est@watson.wustl.edu  
 The library was constructed by Claire Murphy and Dr. James McCarter  
 at Washington University, St. Louis. J2 were provided by Dr.  
 Valerie Williamson of the University of California at Davis  
 (vwilliamson@ucdavis.edu).  
 Seq primer: -40RP from Gibco  
 High quality sequence stop: 424.  
 FEATURES  
 source  
 1. 605  
 /organism="Meloidogyne hapla"  
 /db\_xref="taxon:6305"  
 /clone\_lib="Meloidogyne hapla J2 PAMP1 v1"  
 /dev\_stage="J2"  
 /lab\_host="DH10B"  
 /note="Vector: PAMP1 (Gibco); Site 1: NotI; Site 2: SalI;  
 The library was constructed by Claire Murphy and Dr. James  
 McCarter at Washington University, St. Louis. The cDNA was  
 made by using Dynabead oligo-dT priming (Dynal). PCR based  
 library using a modified protocol from the SMART PCR cDNA  
 Synthesis Kit from Clontech. Directionally cloned into the  
 UDG sites of PAMP1. J2 were provided by Dr. Valerie  
 Williamson of the University of California at Davis  
 (vwilliamson@ucdavis.edu)."  
 BASE COUNT 194 a 98 c 149 g 164 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.02 Length: 605  
 Score: 96.50 Matches: 58  
 Percent Similarity: 39.83% Conservative: 38  
 Best Local Similarity: 24.07% Mismatches: 82  
 Query Match: 4.71% Indels: 63  
 DB: 14 Gaps: 12  
 US-09-889-746-2 (1-400) x B0836176 (1-605)  
 Qy 53 ArgIleThrGlnMetProHisThrThrsValIleTyrGlnGluGlnIleGlnGluGln 72  
 Db 39 AGAGCTGAAAAACATCCCTCACACAGTGGCTGTGATGAAGTCGAACGTTCAAAAGCACTCG 98  
 Qy 73 AlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyVal 92  
 Db 99 ACTACTTCCTCATCCACGAAATGAG-----ACTTTGGGGGTG 137  
 Qy 93 SerSerGlyThrThrsSeranPheGlyInThrMetHisGlyArgGlnValGlnPheLeu 112  
 Db 138 TCCAAATGGGTCTGATATGACATTGGG---ACAGGACAC-----GTCTTTGGCT 185  
 Qy 113 LeuAnGlyValProLeuThrGlySerArgAspIleSerArgGlnLeuAnSerIleAn 132  
 Db 186 TTTCTGCAAAAAACATTTGGAACATCCAAACGAATTAACAACAAATGGGAATCTGTGAAG 245  
 Qy 133 -----ProAnGlnValAlaArgIleGluValLeuSerGlyAlaThrSerIle 148  
 Db 246 GATTATGAAGCCCGGAGAGAGAGTGGCTCAGTAAAGTGGCTAGAGAGAGAGAGATGTT 305  
 Qy 149 TyrGlySerGlyAlaThrGlyLeuIleAnIleValThrLysSerAspLeuGlu 168  
 Db 306 GCCAAAAATTTGATATGAACTCTT----- 332  
 Qy 169 GlnGlnPheGluThrArgIleGlyValHisGlySerLysLeuSer-----SerGlyGly 186  
 Db 333 CCATATGATGAATCCGCGTGTGCAACTTCAATGAAGGATTTGGCAGAAATGGGTATGA 392  
 Qy 189 TyrGlnValGlyGlnSerValAlaGlyValSerGluAnGlyAsnValLeuAla 208  
 Db 393 AAGGTTAT-----ATTAATGCTAGTATGATT--- 419  
 Qy 209 ArgLeuAspValAspTyrArgThrThrThrGlyAlaPheAspAlaAsnGlyValArgIle 226  
 Db 420 -----TTTGATTCGAATCCAAACACCG 443  
 Qy 229 AlaProGluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsnThrAn 246



Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
 http://image.lnl.gov  
 Plate: L1C62472 row: h column: 15  
 High quality sequence stop: 573.  
 Location/Qualifiers

## FEATURES

source

```

1..973
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6280766"
/clone_lib="NIH MGC 113"
/lab_host="DH10B (phage-resistant)"
/note="Organ: spleen; Vector: pOT7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH MGC Library."

```

BASE COUNT 205 a 326 c 241 g 201 t

ORIGIN

Alignment Scores:  
 Pred. No.: 2.23 Length: 973  
 Score: 96.50 Matches: 70  
 Percent Similarity: 33.33% Conservative: 47  
 Best Local Similarity: 19.94% Mismatches: 115  
 Query Match: 4.71% Indels: 119  
 Gaps: 14

US-09-889-746-2 (1-400) x B0712258 (1-973)

```

OY 82 ValMetAlaGlnLeuLeuProSerLeuGlyValSerSerGlyThrThrSerAenPheGly 101
    |||||
DB 77 GATGATACCAAGCTCTCCACACCCCTGCTGTCCTCCGGGGAAGAGCC----- 127
OY 102 GlnThrMetHisGlyArgGlnValGlnPheLeuLeuAenGlyValProLeuThrGlySer 121
    |||||
DB 128 ---ACCTCTCTCGACGAGGCGACGATGATTAAGCAACAACCTGGCCGTGTCCAGCAG 184
OY 122 ArgAspIleSerArgGlnLeuAenSerIleAenProAenGlnValAlaArgIleGluVal 141
    |||||
DB 185 CCA-----CTTGGCCAGGCTCCCGAGGCTCTCATC 214
OY 142 LeuSerGlyAlaThrSerIleTyrgly-----SerGlyAlaThrGlyGly 156
    |||||
DB 215 TATGGCGCATCTCTCGGCTGCTGATTCAGGAGGCTTCAGTGGCAGTGGCTGGG 274
OY 157 LeuIleAenIleValThrIleSerAspLeuGlnGluGlnPheGlnThrArgIleGly 176
    |||||
DB 275 ACAGAGTCTACTCTCTCCATCCAGACCTGCGATCGAAGATT-----GCA 322
OY 177 ValHisGlySerIleLeuSerSerGlnGlyIleGlyTyrglnValGlnSerValAla 196
    |||||
DB 323 GTTATTTTTCACAGATATTAATGACTGGCTCTTACCTTTGGCCAGGG----- 376
OY 197 GlyValSerGlnAenGlnValLeuAlaArgLeuAspValAspTyrrArgThrThrGly 216
    |||||
DB 377 -----ACCAACTGGAAGTCAAA---CGAATGTGGCT 406
OY 217 GlyAlaPheAspAlaAenGlyLysArgIleAlaProGlu-----ProIleGln 232
    |||||
DB 407 -----GCACCATCTGTCTTCATCTTCCCGCATCT 436
OY 233 ThrAspGlnAspSerIleSerLeuSerVal----- 243
DB 437 GATGAGCAGTTGAATCTGAATGCTGTGTGTGCTGCTGCTGAATTAATTCTATCCC 496
OY 244 ---AenThrAsnValAspTrpGlnLeuAspAsp-----LysGlnAenIleAenLeu 259
    |||||
DB 497 AGAGAGGCCAAGATACAGTGAAGTGAATACGCCCTCCATCGGCTAATCCACGAG 556

```

```

OY 260 AlaLeuThrHisTyraAsnAspLysGlnAenThrAspTyraAla----- 273
    |||||
DB 557 AGTGTACACAGAG---CAGACAGCAGACAGCAGCAGCTTACAGCAGACCTGAG 613
OY 274 -----ProAspTyrglyAsnArgLeuAlaValLeuPheGlyGlyLeuProSerLeu 290
    |||||
DB 614 CTGAACAAAGACGACTTACAGAACACACAAAGTCTACGCTTGCAAGCCCCCATCA--- 670
OY 291 AenAlaIleLeuGlyLeuSerLeuSerGlnGlnProLysThrThrIleSerThrPheAen 310
    |||||
DB 671 -----GGGCTGAGCTCGCCGCTACCAAGACTTTAAA----- 706
OY 311 IleAenTyrrHisIleAspAspLeuTrpGlyAenThrIleAenThrAsnAlaTyrrArg 330
    |||||
DB 707 -----GGGAAAGGTTACACCGAACA-----GGC 733
OY 331 ArgGlnLysGlyArgPheTyrrProPheValAlaPro-----PheSerIleAlaLys 347
    |||||
DB 734 CGCCCCCGGGTCCGTACACCCCGCAGAACTCCCCCGGGGTCTTGAACCTTGGCC 793
OY 348 AlaLeuPro----- 350
DB 794 CCTTCTCCCCCAGGGGGGCTTCCCTTATTTGGAGACCCCGCCCGCTATGATTTC 853
OY 351 -----IleLeuGlnSerMetAsnLeuProSerAlaThrLeuAsp 363
DB 854 CGGCGCAAGATCCCGCGGAGATGTTAAATCCCGTCACTCCGCGGCTTTTCTAC 913
OY 364 AlaTyrrThrLysAlaProGlnAlaAlaTyrr 374
DB 914 AGTATGGGAGCTCTCCCGACCGGCGCATTC 946

```

## RESULT 36

BH770394/C

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

INRA

CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France

Tel: 33 1 34 65 25 16

Fax: 33 1 34 65 25 21

Email: sorokine@jouy.inra.fr

best homologue in strain IL1403 is yJae (83%)

Class: shotgun

High quality sequence start: 30

High quality sequence stop: 977.

Location/Qualifiers

source

1..1005

/organism="Lactococcus lactis subsp. cremoris"

/strain="MG1363"

/db\_xref="taxon:1359"

/clone\_lib="MG1363 Random Sequence Tag Library"

/note="Vector: pSGM2; Site\_1: SmaI; Library of

chromosomal fragments of L.lactis strain MG1363 was

prepared by partial AluI digestion or by sonication."

BASE COUNT 258 a 206 c 186 g 355 t

ORIGIN

Alignment Scores:

Pred. No.: 2.35 Length: 1005

Score: 96.50 Matches: 65  
Percent Similarity: 44.49% Conservative: 56  
Best Local Similarity: 23.90% Mismatches: 84  
Query Match: 4.71% Indels: 67  
DB: 17 Gaps: 18

US-09-889-746-2 (1-400) x BH770394 (1-1005)

Qy 17 ValAlaValThrGlnGlnLeuTyrAlaGlnProAenGluSerLeuProThrValGluLeu 36  
|||||  
Db 929 GTCAGTGGAACTCAACAATTA---GCTAGGGAACCCAACTTG----- 888

Qy 37 GluProValValLeuThrLeuAspSerGlyMetAlaLeuAlaAsnArgileThrGln 56  
|||||  
Db 887 -----GCTGATAAATCT---AAAGATCTGTACGACGGAATTTCCACAG 849

Qy 57 MetProHisThrThrLysValIleTyrGluGluGlnIleGlnGluAlaThrGlySer 76  
|||||  
Db 848 ATTTCTCAAGTACAGACGGGTGACCCAAATGCAATG-----GGTGTG 804

Qy 77 ArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyValSerGlyThr 96  
|||||  
Db 803 CAACAACATCCGAGGCTTAAGTCAATGGCTCAAAAACAACTTTATCCGAAACTCAA 744

Qy 97 ThrSerAsnPhcGlyGlnThrMetHisGlyArgGlnValGlnPheLeuLeuAsnGlyVal 116  
|||||  
Db 743 CAGCAAAATATTAGTCTGTGCAAAATGCT----- 714

Qy 117 ProLeuThrGlySerArgAspIleSerArgGlnLeu-----AsnSerIleAsnProAsn 134  
|||||  
Db 713 ---CTTACT-----GAATTAATCAAGAGTTACAAAATAATACGATTGACCCAAAT 666

Qy 135 GlnValAlaArgileGluValLeuSerGlyAlaThrSerIleTyrGlySerGlyAlaThr 154  
|||||  
Db 665 TTGGCTGCTAATGTCAA-----AAAACCTGGAAGGTCTTGGAGTACG 621

Qy 155 ---GlyGlyLeuLeuAsn-----IleValThrLysSerAspLeuGluGlu---Glu 169  
|||||  
Db 620 GTTACAAATTTAGCTAATGATCAAGTGGCAGCCTCTAAGGCAGCAGTGTGAAAGTACTGAG 561

Qy 170 GlnPheGluThr-----ArgIleGlyValHisGlySerLysLeuSerSer 184  
|||||  
Db 560 ACTTTTAAACCTTTAACCAAAAGTCAGAAACAGATATAGAAATGCATTAGATAATCT 501

Qy 185 GluGlyIleGly-----TyrGlnValGlyGlnSerValAlaGlyVal 198  
|||||  
Db 500 GTTGTGCGGAGCGGTTCAAGCTGATTTAGTTTCAATTAATGAGAATATTACAGGGGTC 441

Qy 199 SerGluAsn---GlyAsnValLeuAlaArgLeuAspValAspTyrArgThrThrGlyGly 217  
|||||  
Db 440 AAAAAACAGTCTTACAAATCTACTACCAATGCTT-----GGA 405

Qy 218 AlaPheAspAlaAsnGlyLysArgileAlaProGluProAlaGlnThrAspLysGlnAsp 237  
|||||  
Db 404 AAAATT---GCTGATCTTAAACATTAGTTCCAGGGGCAAGTCNAACAATTTACTGACCTT 348

Qy 238 SerLysSerLeu---SerValAsnThrAsnValAspTrpGln---LeuAspAspLysGln 255  
|||||  
Db 347 TCTACAGGTTTAAATAGTGTGATAGTAGCTTAAATAGTCAAAATTTTACCAGGAATGAAT 288

Qy 256 AsnIleAsnLeuAlaLeuThrHisTyrAsnAspLys 267  
|||||  
Db 287 AAACTAGACAGTGGTTTGACAACTTCAATAGTCAA 252

RESULT 37  
BE015300/c 509 bp mRNA linear EST 02-OCT-2000  
LOCUS BE015300 Manduca sexta female antennae Uni-ZAP XR library Manduca  
DEFINITION sexta cDNA clone pMsfab22 5' similar to cuticle protein, mRNA  
sequence.  
ACCESSION BE015300  
VERSION BE015300.1 GI:8276350  
KEYWORDS EST.

SOURCE tobacco hornworm.  
ORGANISM Manduca sexta  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;  
Sphingioidea; Sphingidae; Sphinginae; Manduca.  
REFERENCE 1 (bases 1 to 509)  
AUTHORS Robertson,H.M., Waiden,K.K.O., Brakebill,C.M., Mostafavipour,P.,  
Schmidt,L. and Nardi,J.B.  
TITLE Expressed sequence tags from the antennae of female Manduca sexta  
moths  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robertson HM  
Department of Entomology  
University of Illinois at Urbana-Champaign  
505 S. Goodwin, Urbana, IL 61801, USA  
Tel: 217 333-0489  
Fax: 217 244 3499  
Email: hughrobes@uiuc.edu  
Insert Length: 700 Std Error: 0.00  
Seq primer: T3  
High quality sequence stop: 450.  
FEATURES  
source  
1..509  
/organism="Manduca sexta"  
/db\_xref="taxon:7130"  
/clone="pMsfab22"  
/clone\_lib="Manduca sexta female antennae Uni-ZAP XR  
library"  
/sex="female"  
/dev\_stage="newly eclosed adults and pharate adults"  
/lab\_host="XLI Blue MRF" and SOLR"  
/notes="Organ: antennae; Vector: Uni-ZAP XR; Site 1: EcoRI;  
Site 2: XhoI; The library was prepared by Stratagene using  
oligo-T priming and unidirectional cloning with an  
adapter at the 5' end (G3CAGCAG) following the EcoRI  
site. The mRNA was prepared from antennae of late pupal  
and newly eclosed female moths. Clones were subcloned in  
vivo in mass into pBluescript maintained in SOLR cells for  
DNA sequencing."  
BASE COUNT 74 a 208 c 140 g 87 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 0.878 Length: 509  
Score: 96.00 Matches: 45  
Percent Similarity: 37.84% Conservative: 25  
Best Local Similarity: 24.32% Mismatches: 73  
Query Match: 4.69% Indels: 42  
DB: 10 Gaps: 8  
US-09-889-746-2 (1-400) x BE015300 (1-509)  
Qy 79 LeuAlaAspValMetAlaGlnLeuIleProSerLeu---GlyValSerSerGlyThrThr 97  
Db 487 ATGGAGTCCGTAAAGCCCGGAGCCCATGGACTCCGTACGGCGGTAGCGAGCGCGGAC 428  
Qy 98 -----SerAsnPhcGlyGlnThrMetHisGlyArgGlnValGlnPheLeuLeuAsnGly 115  
Db 427 GAAGTGACCACGCTAGGGGCGAGCATAAGGGCGCGGTAGTGTAGCTCCGTAAGGGCGCG 368  
Qy 116 ValProLeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAsnProAsnGln 135  
Db 367 CAGGGCAGCGTGGCAGCGCGGAC-----ACCGCTGGCGCGCGCGCG 326  
Qy 136 ValAlaArgileGluValLeuSerGlyAlaThrSerIleTyrGlySerGlyAlaThr--- 154  
Db 325 GTGGGA-----CTGAGCGCGGCAACAGCGTTTGGTGGCGCGCAAGACCGTT 278  
Qy 155 -----GlyGlyLeuLeuAsnIleValThrLysSerAspLeuGluGluGlnPheGlu 172  
Db 277 CCAGTAGCGGCTCTGTGAGCGTGGCGCTGGCTCGGCTACTCGCGGGAAGCTTCAGC 218  
Qy 173 ThrArgileGlyValHisGlySerLysLeuSerSerGluGlyIleGlyTyrGlnValGly 192



	DB	217	GTTGAAGTCTG-----GGCAGCGGCTTCACGCAGCGCGCGCCCTGACGTATGACAGC	167
OY	193	GinservAla1a6g1ValserGelunenglyanvalLeuAlaArgLeuAspValaspyr	212	
DB	166	GGCAGCGTGTC-----CTGGCGGCCACCGA	140	
OY	213	ArgThrThrglVglValaphe-----Asp	220	
DB	139	CGCGTCGATGGCGGGGGCGCTGGCGCGCACGTTCCACCGCTAAAGCTAGGAGAGAT	80	
OY	221	AlaennglyLysArgllLeAlProclupPoaIagIntHraspleGlnaspSerlysser	240	
DB	79	GGCGGTAGCGAGCATGACCAGATGGCGACGCAAGAACCATCAAGTAATCTCATGTCGA	20	
OY	241	LeuSerValAsnThr	245	
DB	19	CTTCTTGTCACACT	5	
RESULT	38			
LOCUS	B0836774	593 bp	mRNA	linear
DEFINITION	B0836774	rfc4a08.v1	Meloidogyne hapla J2 PAMP1 v1	Meloidogyne hapla CDNA 5'
VERSION	B0836774			similar to TR:Q9Y419 Q9Y419 TYROSINE PHOSPHATASE IA-2BETA ; mRNA
KEYWORDS	B0836774.1	GI:22141092		
SOURCE	EST.			
ORGANISM	Meloidogyne hapla.			
REFERENCE	Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;			
AUTHORS	Tylenchoidea; Heteroderidae; Meloidogyninae; Meloidogyne.			
	1 (bases 1 to 593)			
	McCarteer,J., Clifton,S., Chiappelli,B., Pape,D., Martin,T., Wylie,T.,			
	Dance,M., Marx,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,			
	Gibbons,M., Rittler,E., Bennett,J., Franklin,C., Tsagarisbivill,R.,			
	Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe			
	,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,			
	Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and			
	Wilson,R.			
TITLE	The Washington Univ. Nematode EST Project, 1999			
JOURNAL	Unpublished (1999)			
COMMENT	Contact: McCartney The Washington Univ. Nematode EST Project, 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel.: 314 286 1800 Fax: 314 286 1810 Email: est@watscn.wustl.edu The library was constructed by Claire Murphy and Dr. James McCartney at Washington University, St. Louis. J2 were provided by Dr. Valerie Williamson of the University of California at Davis (vwilliamson@ucdavis.edu). Seq primer: -40RP from Gibco High quality sequence stop: 417.  Location/Qualifiers 1..593 /organism="Meloidogyne hapla" /db_xref="taxon:6305" /clone_lib="Meloidogyne hapla J2 PAMP1 v1" /dev_stages="J2" /lab_host="MDH10B" /note="Vector: pAMP1 (Gibco); Site 1: NotI; Site 2: SalI; The library was constructed by Claire Murphy and Dr. James McCartney at Washington University, St. Louis. The cDNA was made by using Dynabead Oligo-dT priming (Dyna1). PCR based library using a modified protocol from the SMART PCR cDNA Synthesis Kit from Clontech. Directionally cloned into the UDG sites of pAMP1. J2 were provided by Dr. Valerie Williamson of the University of California at Davis (vwilliamson@ucdavis.edu)."			
BASE COUNT	190 a	96 C	146 G	161 T
ORIGIN				

Alignment Scores:		Pred. No.:	1.13	Length:	593
Score:	96.00	Matches:	56		
Percent Similarity:	40.17%	Conservative:	38		
Best local Similarity:	23.93%	Mismatches:	84		
Query Match:	4.69%	Indels:	56		
DB:	14	Gaps:	11		

  

US-09-889-746-2 (1-400) x B0836774 (1-593)	
Qy	53 ArgIleThrgInmeProhIstThrThryValIleTyrgIuGlInIleGIngluIn 72
Db	39 AGAGCTGAACATCCCTCACACAGTTCCTGTGATCGAAGTCACAGCTTCAAGACCTCG 98
Qy	73 AlathrgIySerArGInleuAlaAspValMetAlaGInleuIleProSerleuGlyVal 92
Db	99 AGTACTCTGTCATTCACGAAATATGAG-----ACTTTGGGGTGG 137
Qy	93 SerSerGlyThrThrSerAsnPhedIyGInThrMetHISGlyArGInValIGlnPheIeu 112
Db	138 TCCATGAGGTCTGATATGACATTGGC--ACAGACAC--GTCCTTTGGCTTTCTT 191
Qy	113 leuAenGlyValProleuThrgIySerArGAspIleSerArGInleuAenSerIleAsn 132
Db	192 GTTAA-----CATTTGGAACATCCAAACGAATTAACAAACATGGAGATCTGTAG 245
Qy	133 -----ProAenGInValAlaArgIleGluValleuSerGlyAlaThrSerIle 148
Db	246 GATTATGAGCGCCGCGAAGAGTGGCTCAGTAAAGGCTAGAGAGACAGACGATGTT 305
Qy	149 TyrgIySerGlyAlaThrgIyGlyleuIleAenIleValThrIySerAspIleuGlu 168
Db	306 GCCAAATTTTGATTAATGCAACTTT----- 332
Qy	169 GluInIleGluThrArgIleGlyValHISGlySerIyIleuSerSerGluGlyIleGly 188
Db	333 CCATATGATGATGATCCGCTGTCACACTTCATGAAGCGATTGGCAGAAATGGTGATGA 392
Qy	189 TyrgInValIGlyGInSerValAlaGlyAlaSerIuAenGlyAenValleuAlaArgIeu 208
Db	393 -----AAGGTTTATATTAAATGCTAGTAA 416
Qy	209 AspValAspTyArgThrThrgIyGlyAlaPheAspAlaAenGlyIyAsArgIleAlaPro 228
Db	417 -----ATTTTATTGGAATCCAGAACACCGCTTAT 449
Qy	229 GluProAlaGInThrAspIyGlnAspSerIySerIeuserValAsnThrAsnValAsp 248
Db	450 ATTGCGCGCCCAACACCTCTAGAGCAAACTGCTCTCAATT----- 491
Qy	249 TrpGInleuAspAspIyGlnIleAsnIleAenleuAlaIeuThrHISTyArgAspIyGln 268
Db	492 TGGCAAAATGATTGGGAACAGGAAATGGGTTA--ATCGTCAATTGTGTGATTAAGA 548
Qy	269 AspThr-----AspTyAlaProAspTyArgIyAsnArgIeu 280
Db	549 GACATTTACTTTAAATATTGGCAGAGAGAGAGAAAGCTG 590

  

RESULT 39	
LOCUS	A1133960
DEFINITION	A1133960 749 bp mRNA linear EST 23-APR-2001
ACCESSION	GI11267.5
VERSION	GI11267.5
KEYWORDS	melanogaster cDNA clone GI11267.5 similar to ACDX: P8and005712
SOURCE	EST.
ORGANISM	fruit fly.
	Drosophila melanogaster
	Eukaryote; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
	Ephydroidea; Drosophilidae; Drosophila.



Alignment Scores:  
 Pred. No.: 2.4 Length: 863  
 Score: 95.50 Matches: 59  
 Percent Similarity: 38.34% Conservative: 38  
 Best Local Similarity: 23.32% Mismatches: 91  
 Query Match: 4.67% Indels: 65  
 DB: 17 Gaps: 14

US-09-889-746-2 (1-400) x BH396248 (1-863)

QY 95 GYTHrThSerAspPheGly---GlnThrMetHisGlyArgGlnValGlnPheLeu 113  
 |||  
 DB 77 GGTTCAGATTATTAATTTCTCTATCAACGGTATGTCGGTAAAGCGGTGAATTTTATA 136  
 QY 114 AangIyValProLeu-----ThrglySerArgAspIleSerArgGlnLeuAsnSerile 131  
 |||  
 DB 137 GACGGAGTCTCTTGGAAATGTTGGCAAAAGGTGTTTC-----CTGAGCACACG 187  
 QY 132 AenProAsnGlnValAlaArgIleGlnValLeuSerGlyAlaThrSerIleTy---Gly 150  
 |||  
 DB 188 CCTGTAATATGACCGACCGATGATTTAAAGGTGTGTCCTCGTTCATCTGAGT 247  
 QY 151 SerGlyAlaThrglyGlyLeuIleAsnIleValThrLysSer----- 164  
 |||  
 DB 248 ACCGATGCTATGCGGGAGCTGTAAATATTATTAAGTCTTCGCGCAATAAACTATCTG 307  
 QY 165 -----AspLeuGlnGluGlnGlnPheGlnThr---ArgIleGlyValHisGlySer--- 180  
 |||  
 DB 308 GATGTCGGATGATGATGTCGGATCTTTATATACGACCGATCACTGATGATGACGATC 367  
 QY 181 LylLeuSerSerGlnGlyIleGlyTyGlnValGlyGlnSerValAlaGlyValSerGlu 200  
 |||  
 DB 368 AAAGTATGATAAACCGGATTTATCTCGCGATC---AATAGTTTATTAACCATTCGAT 424  
 QY 201 AenGlyAsnValLeu----- 205  
 |||  
 DB 425 AACAACTATATCATGAGATATGAAGATATGATGATCCGCAAAATGATATGATTA 484  
 QY 206 -----AlaArgLeuAspValAspIleArgTrhThrglyValAspAspAlaAsn 222  
 |||  
 DB 485 AGAATGTAATAACGTTTCAACGACGTTATCACTGCTGTTTGGATGACAGAACCGGA 544  
 QY 223 GlyLeuArg-----IleAlaProGlnProAlaGlnThrAsp 234  
 |||  
 DB 545 GTAGAAATATAAAATGCGGACATATTTTGTATGATGATCAACAGCTGTTTCCAGC 604  
 QY 235 LysGlnAspSerLysSerLeuSerValAsnThrAsnValAspTrpGlnLeu----- 251  
 |||  
 DB 605 AAACAG-----GTACAAACAGATCCAAATCAGAGATGTTTATATGAT 646  
 QY 252 AspAspLysGlnAsnIleAsnLeuAlaLeuThrHisTyAsnAspLysGlnAspThrAsp 271  
 |||  
 DB 647 GCGCGTGAAGCAGAT-----GACACGGCTCTAAT 676  
 QY 272 TyraIaProAspTyrglyValAsnArgLeuAlaValLeuPheGlyGlyLysProSerLeuAsn 291  
 |||  
 DB 677 TATTTATGAATAT-----AAAAAGCCCAATCTTTTAAACGACCGGTATGATCTTAT 730  
 QY 292 AlaIleLysGlyLeuSerLeuSerGlnGlnProLysThr 304  
 |||  
 DB 731 GTTATGACGTTTTCCTCAAAAGTGTCCAGAAAGCTACT 769

RESULT 41  
 BH370350 590 bp DNA linear GSS 10-DEC-2001  
 LOCUS BH370350  
 DEFINITION AG-ND-140C15.TR ND-TAM Anopheles gambiae genomic clone AG-ND-140C15  
 , DNA sequence.  
 ACCESSION BH370350  
 VERSION BH370350.1 GI:17316475  
 KEYWORDS GSS.  
 SOURCE African malaria mosquito.  
 ORGANISM Anopheles gambiae  
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
 Anopheles.  
 1 (bases 1 to 590)  
 REFERENCES  
 Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.  
 TITLE Direct Submission of BAC-end sequences from Anopheles gambiae  
 JOURNAL Unpublished (2001)  
 COMMENT  
 Other\_GSSs: AG-ND-140C15.TP  
 Contact: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0208  
 Fax: 301 838 3543  
 Email: b.loftus@igf.org

This clone is from an A. gambiae BAC library (ND-TAM) provided by  
 F.H. Collins and sequenced by The Institute for Genomic Research  
 (TIGR). The BAC library was generated from A. gambiae PEST strain  
 DNA. All DNA was extracted from newly hatched first instar larvae  
 to minimize the inclusion of DNA from microorganisms that inhabit  
 the gut. The DNA is derived from mixed sexes of larvae. The BAC  
 library was constructed at Texas A&M University BAC Center  
 University, College Station, Texas 77843-2123, USA using a HindIII  
 partial digest.  
 Seq primer: M13 Rev  
 Class: BAC ends.

FEATURES  
 source location/Qualifiers  
 1..590  
 /organism="Anopheles gambiae"  
 /strain="PEST"  
 /db\_xref="taxon:7165"  
 /clone="AG-ND-140C15"  
 /clone\_id="ND-TAM"  
 /note="Vector: pECBAC1; Site\_1: HindIII"

BASE COUNT 219 a 86 c 129 g 156 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1.47 Length: 590  
 Score: 95.00 Matches: 40  
 Percent Similarity: 44.57% Conservative: 42  
 Best Local Similarity: 21.74% Mismatches: 57  
 Query Match: 4.64% Indels: 45  
 DB: 17 Gaps: 8

US-09-889-746-2 (1-400) x BH370350 (1-590)

QY 91 GlyValSerSerGlyThrThrSerAsnPheGlyGlnThrMetHisGlyArgGlnValGln 110  
 |||  
 DB 118 GGTATATCTATGTTGTTGAAGTAAAGTAACTATGATGATTAATGACGAATGTACAG 177  
 QY 111 PheLeuLeuAsnGlyValProLeuThrglySerArgAspIleSerArgGlnLeuAsnSer 130  
 |||  
 DB 178 -----ATGTCGGTCT--GAACATATGAATTAATTTGAAGTCG 213  
 QY 131 IleAsnProAsnGlnValAlaArgIleGlnValLeuSerGlyAlaThrSerIleTyrgly 150  
 |||  
 DB 214 GTAAATCTGAATAATATGTCGCAAAATAGAAATTAACAAACCTCCGGCTAAATATGAT 273  
 QY 151 SerGlyAlaThrglyGlyLeuIleAsnIleValThrLysSerAspLeuGlnGlnGln 170  
 |||  
 DB 274 GCCCAGGAAATAGTGTATCTTATTAATTTGAGAGAAAT----- 318  
 QY 171 PheGlnThrArgIleGlyValHisGlySer-----LysLeuSerSer 184  
 |||  
 DB 319 -----TCCAAATCTGGGCTGGAGTGAATCTGACTACGCAATATATAGAAAGTCACGG 372  
 QY 185 GlnGlyIleGlyTyrglyValGlnSerValAlaGlyValSerGlnAsnGlyVal 204  
 |||  
 DB 373 GATGATTTGTAATAAATGATCGGCTGAATAT-----CATCTGGTAAAGATC 420  
 QY 205 LeuAlaArgLeuAspValAspIleArgTrhThrglyGlyAlaPheAspAlaAsnGlyLys 224  
 |||  
 DB 421 AGCAGCTCATTAAGTAAAGCATTAATGATACC-----AAA 456

```

QY 225 ArgIleAlaProGluProLa-----GlnThrAspLysGln-AspSe 238
Db 457 AAGAATTCTTCAGAGCGGAATGGAATATTGGCAATAAAGGACTGAAAGCCTTGACAG 516
QY 238 rlys-----SerLeuSerValIleThrAsnValAspTrpGlnLeuAspAe 253
Db 517 AGAGTAGACATGCCTTAATGGTCTTGCAATTATAGTCTGGATTATAAATATACGA 576

QY 253 pLysGlnAsn 256
Db 577 AAGTCTAAT 586

RESULT 42
CNS06PDW/c
LOCUS
DEFINITION
T7 end of clone AV0AA013G05 of library AV0AA from strain CBS 379 of
Saccharomyces exiguus, genomic survey sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Saccharomyces exiguus.
Saccharomyces exiguus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
1 (bases 1 to 979)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Portier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
20584711
11152876
2 (bases 1 to 979)
Bon,E., Neuveglise,C., Lepingle,A., Wincker,P., Artiguenave,F.,
Gaillardin,C. and Casaregola,S.
Genomic exploration of the hemiascomycetous yeasts: 6.
Saccharomyces exiguus
FEBS Lett. 487 (1), 42-46 (2000)
20584716
11152881
3 (bases 1 to 979)
Genoscope.
Direct Submission
Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
Location/Qualifiers
1..979
/organism="Saccharomyces exiguus"
/strain="CBS 379"
/db_xref="taxon:34358"
/clone="AV0AA013G05"
/clone_lib="AV0AA"
/note="end : T7"
BASE COUNT 226 a 205 c 217 g 331 t
ORIGIN
Alignment Scores:

```

```

Pred. No.: 3.38 Length: 979
Score: 95.00 Matches: 63
Percent Similarity: 37.84% Conservative: 35
Best Local Similarity: 24.32% Mismatches: 107
Query Match: 4.64% Indels: 55
DB: 17 Gaps: 11

US-09-889-746-2 (1-400) x CNS06PDW (1-979)

QY 31 LeuProThrValGluLeu---GluProValValIleThrIleAspLysSerGlyMetAla 49
Db 860 CTACCACTACAGGAACCTGGTGCACCGATGGTAATGGTTCTCAAA-----CTACAA 810
QY 50 LeuAlaAsnArgIleThrGlnMetProHisThrLysValIleTyrGluGlnIle 69
Db 809 CTACAGGAACCTGGTGCACCTGATGGTAATGGTTCTCAAACTACCACTACAGGACCGGTG 750
QY 70 GlnGluGlnAlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSer 89
Db 749 CAACT-GATGGTAATGGTTCTCAAACTACACCGGATGGTAATAGTGCTCAAACTACCCT 691
QY 90 LeuGlyVal-----SerSerGlyThrThrSerAsnPheGlyGln 102
Db 690 ACAGGAACCTGGTGCACCGGATGGTAATGGTTCTCAAACTACCACTACAGGAACCTGGTGA 631
QY 103 ThrMetHisGlyArgGlnValGlnPheLeuLeuAsnGlyValProLeuThrGlySerArg 122
Db 630 ACT---GATGGTAATGGTTCTCAAACTACCACTACAGGA-----ACCGGTGCAACC 583
QY 123 AspileSerArgGlnLeuAsnSerIleAsnProAsnGlnValAlaArgIleGluValLeu 142
Db 582 GATGGTAACGGCTCTTTAACTACAACCTGATGATAATAGTGCTCAAACTACCACTACAGA 523
QY 143 SerGlyVala-----ThrSerIleTyrGlySerGlyAlaThrGly 155
Db 522 ACTGGTGCACCGGATGGTAACCGCTCTTTAAACACCACTACAGGAACCTGGTGCACCTGAT 463
QY 156 GlyLeuIleAsnIleValThrLysSerAspLeuGluGluGlnPheGluThrArgIle 175
Db 462 GGTAACGGCTCTTTAACTACAACCTGATGATAATAGTGCTCAAACTACCACTACAGGAAC 403
QY 176 GlyVal-----HisGlySerLysLeuSerSerGlyGlyIleGlyTyrGlnValGly 192
Db 402 GGTGCAACTGATGGTAATGGTTCTCAAACTACACCGGATGGT----- 361
QY 193 GlnSerValAlaGlyValSerGluAsnGlyAsnValLeuAlaArgLeuAspValAspTyr 212
Db 360 -----AATGGTTCTCTTAACCAACC----- 343
QY 213 ArgThrThrGly---GlyAlaPheAspAlaAsnGlyLysArgIleAlaProGluProAla 231
Db 342 ---ACTACAGGAACCTGGTGCACCGGATGGTAATGGT-----TCT 307
QY 232 GlnThrAspLysGlnAspSerLysSerLysSerValAsnThrAsnValAspTrpGlnLeu 251
Db 306 CAAAGTACCAACTGATGGTAATGGTTCTTTAACTGCAACTGATGGTAACAGTGTCTCAACT 247
QY 252 AspAspLysGlnAsnIleAsnLeuAlaLeuThrHisTyrAsnAspLysGlnAspThr 270
Db 246 ACAACTGATGCAATGGTTCTTCTTAACCGGATGGTAACAGTGTCTCAAACTACC 190

RESULT 43
BQ836244
LOCUS
DEFINITION
rf43f09, Y1 Meloidogyne hapla J2 PAMPI v1 Meloidogyne hapla cDNA 5',
similar to TR:Q9Y4I9 Q9Y4I9 TYROSINE PHOSPHATASE 1A-2BETA ;, mRNA
sequence.
ACCESSION BQ836244
VERSION BQ836244.1 GI:22140558
KEYWORDS EST.
SOURCE Meloidogyne hapla.
ORGANISM Meloidogyne hapla
Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;

```



McCarter at Washington University, St. Louis. The cDNA was made by using Dynabead oligo-dT priming (Dyna). PCR based library using a modified protocol from the SMART PCR cDNA Synthesis Kit from Clontech. Directionally cloned into the UDG sites of pAMP1. J2 were provided by Dr. Valerie Williamson of the University of California at Davis (vwilliamson@ucdavis.edu).

BASE COUNT 193 a 101 c 151 g 166 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.78 Length: 611  
Score: 94.50 Matches: 57  
Percent Similarity: 39.33% Conservative: 37  
Best Local Similarity: 23.85% Mismatches: 86  
Query Match: 4.62% Indels: 59  
DB: 14 Gaps: 11

US-09-889-746-2 (1-400) x BQ935870 (1-611)

Qy 53 ArgileThrGlnMetProHisThrThrLysValIleTyrGluGlnIleGlnGluGln 72  
Db 39 AGAGCTGAAACATCCTCACAGTTGCTGATCGAAGGTCAACGTTCAAGCACTCG 98  
Qy 73 AlaThrGlySerArgGlnLeuAlaAspValMetAlaGlnLeuIleProSerLeuGlyVal 92  
Db 99 AGTACTTCGTCTATCCACGAAATGAG-----ACTTTTGGGGTG 137  
Qy 93 SerSerGlyThrThrSerAsnProHisGlnThrMetHisGlyArgGlnValGlnPheLeu 112  
Db 138 TCCATGGGTCGTATGACATTGGG-----ACAGGAC-----GTCCTTTGGCT 185  
Qy 113 LeuAsnGlyValProLeuThrGlySerArgAspIleSerArgGlnLeuAsnSerIleAsn 132  
Db 186 TTTCTGCAAAACATTTGGACATCCAAACGAAATTAACAACAATGGGAATCTGTGAAG 245  
Qy 133 -----ProAsnGlnValAlaArgIleGluValLeuSerGlyAlaThrSerIle 148  
Db 246 GATTATGAAGCGCGGGAAGTGGCTCAGTGTAAAGTGGCTAGAGAGAGAGAGATGTT 305  
Qy 149 TyrGlySerGlyAlaThrGlyGlyLeuIleAsnIleValThrLysSerAspLeuGlu 168  
Db 306 GCCAAATTTTGATGAACTCTT----- 332  
Qy 169 GluGlnPheGluThrArgIleGlyValHisGlySerLysLeuSerSerGluGlyIleGly 188  
Db 333 CCATATGATGAATCGCTGCGCACTTCATGAAAGCGATTTGGCAGAAATTTGGGTATGA 392  
Qy 189 TyrGlnValGlyGlnSerValAlaGlyValSerGluAsnGlyAsnValLeuAlaArgLeu 208  
Db 393 -----AAGGGTTATATATATGCTAGTATG 416  
Qy 209 AspValAspTyrArgThrThrGlyGlyAlaPheAspAlaAsnGlyLysArgIleAlaPro 228  
Db 417 -----ATTTTTGATTCGAATCCAGAACCAACCGCTTAT 449  
Qy 229 GluProAlaGlnThrAspLysGlnAspSerLysSerLeuSerValAsnThrAsnValAsp 248  
Db 450 ATTGCGCCCAACACCTCTAGACAACTGCTTCTCAATTT----- 491  
Qy 249 TrpGlnLeuAspAspLysGlnAsnIleAsnLeuAlaLeuThrHisTyrAsnAspLysGln 268  
Db 492 TGGCAATGATTTGGGAACAGGGAATTTGGGTTA---ATCGTCAATTTGTGTGATAAAGGA 548  
Qy 269 AspThr-----AspTyrAlaProAspTyrGlyAsnArgLeuAlaValLeuPheGly 285  
Db 549 GACAATTTACTTAAATATTGGCCAGAGGAGGATGAAG-----CTGTTTGGT 596

RESULT 45  
BJ423365/c 649 bp mRNA linear EST 11-MAR-2002  
LOCUS  
DEFINITION  
BU423365 Dictyostelium discoideum cDNA library, VF Dictyostelium  
discoideum cDNA clone ddv48019 5', mRNA sequence.

ACCESSION BJ423365  
VERSION BJ423365.1 GI:19340072  
KEYWORDS EST.  
SOURCE Dictyostelium discoideum.  
ORGANISM Dictyostelium discoideum  
Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.  
REFERENCE 1 (bases 1 to 649)  
AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.  
TITLE Full length cDNA of Dictyostelium discoideum at the vegetative stage  
JOURNAL Unpublished (2002)  
COMMENT Contact: Tadaasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.  
FEATURES  
Location/Qualifiers  
1..649  
/organism="Dictyostelium discoideum"  
/strain="AX4"  
/db\_xref="taxon:44689"  
/clone="ddv48019"  
/clone\_lib="Dictyostelium discoideum cDNA library, VF"  
/sex="mat A"  
/dev\_stage="Growth phase"  
BASE COUNT 184 a 97 c 113 g 255 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.96 Length: 649  
Score: 94.50 Matches: 39  
Percent Similarity: 40.34% Conservative: 32  
Best Local Similarity: 22.16% Mismatches: 90  
Query Match: 4.62% Indels: 15  
DB: 13 Gaps: 6

US-09-889-746-2 (1-400) x BJ423365 (1-649)

Qy 137 AlaArgIleGluValLeuSerGlyAlaThrSerIleTyrGlySerGlyAlaThrGlyGly 156  
Db 637 GCCTCAATACAAATTTTAACTAGTCGACATCATCTAGTGGGTCAAGTACAACTGGTGA 578  
Qy 157 LeuIleAsnIleValThrLysSerAspLeuGluGluGlnPheGluThrArgIleGly 176  
Db 577 -----TCAGCCACATCTTCAACACCATCTTCAACATCATAGTAATACTGGA 527  
Qy 177 ---ValHisGlySerLysLeuSerSerGluGlyIleGlyTyrGlnValGlnSerVal 195  
Db 526 TCAGTACATGGAATAAAGCTAGTCATATGTCAGGTTGGGATTCAGTCCAACTTT 467  
Qy 196 AlaGlyValSerGluAsnGlyAsnValLeuAlaArgLeuAspValAspTyrArgThrThr 215  
Db 466 TCAGCAACCAATGTAGAT---AACGTTATG-----GAAATGCAAATGCT 425  
Qy 216 GlyGlyAlaPheAspAlaAsnGlyLysArgIleAlaProGluProAlaGlnThrAspLys 235  
Db 424 GGAATGAAGTAAATGAAACCAATGAATACTTTGAAGACCAAAATGGACAAACCAATGAA 365  
Qy 236 Gln-----AspSerLysSerLeuSerValAsnThrAsnValAspThrGlnLeu 251  
Db 364 ACAATCAAAATTTATATCAAAACATAGAAATTAACAACACACAGAGATTGG---TTA 308  
Qy 252 AspAspLysGlnAsnIleAsnLeuAlaLeuThrHisTyrAsnAspLysGlnAspThrAsp 271  
Db 307 TCACACACACAGGTTATTTGGTATATAGTATCTCCACAGACGGTTCATTAGAAATTCAT 248  
Qy 272 TyrAlaProAspTyrGlyAsnArgLeuAlaValLeuPheGlyGlyLysProSerLeuAsn 291  
Db 247 ATCCAGAAAACTTATATGCTATGAAGACAAATATTTGACCAATCTGCTAGTAGGT 188  
Qy 292 AlaIleLysGlyLeuSerLeuSerGluGlnProLysThrThrLysSer 307

Thu Jan 2 15:16:41 2003

us-09-889-746-2.rst

Page 35

Db 187 CGTTCTCGAGATTCTTATGATGATACCCCTAAAAATCCATATAGT 140

Search completed: December 25, 2002, 21:58:59  
Job time : 1976 secs

**THIS PAGE BLANK (USPTO)**